

THE
PROCEEDINGS
OF
THE PHYSICAL SOCIETY

FROM JANUARY 1945 TO NOVEMBER 1945

VOLUME 57

Published by
THE PHYSICAL SOCIETY
1 Lowther Gardens, Exhibition Road,
London S.W.7

Printed by
TAYLOR AND FRANCIS, LTD.,
RED LION COURT, FLEET STREET, LONDON E.C. 4

OFFICERS AND COUNCIL, 1944-45

PRESIDENT

D. BRUNT, M A , Sc D , F R S

VICE-PRESIDENTS

who have filled the Office of President

C H. LEES, D.Sc., F.R.S. (1918-20)
Sir RANK E. SMITH, G.C.B , G.B.E., D.Sc., LL.D., F.R.S. (1924-26)
Sir OWEN W. RICHARDSON, M.A , D.Sc , F.R.S. (1926-28)
W. H. ECCLES, D.Sc., M.I.E.E., F.R.S (1928-30)
A. O. RANKINE, O.B.E , D.Sc , F.R.S. (1932-34)
The Right Hon. LORD RAYLEIGH, M.A , Sc D., LL D , F.R.S. (1934-36)
T SMITH, M A., F.R.S (1936-38)
ALLAN FERGUSON, M.A., D.Sc. (1938-41)
Sir CHARLES G. DARWIN, K.B.E., M.C., M.A., Sc.D., F.R.S (1941-43)
E N. da C ANDRADE, Ph.D , D Sc., F.R.S (1943-45)

VICE-PRESIDENTS

H. T. FLINT, D.Sc., Ph.D.	Sir EDWARD V. APPLETON, K.C.B
N. F. MOTT, M.A., F.R.S.	M.A., D.Sc., LL.D., F.R.S.
	S CHAPMAN, M A , D Sc , F R S

HONORARY SECRETARIES

W. JEVONS, D.Sc., Ph.D. (Business) J. H. AWBERRY, B.A., B.Sc. (Papers)

HONORARY FOREIGN SECRETARY

E N da C. ANDRADE, Ph D , D.Sc., F.R.S.

HONORARY TREASURER

C. C. PATERSON, O.B.E , D.Sc., M.I.E.E., F.R.S.

HONORARY LIBRARIAN

L. C. MARTIN, D.Sc.

ORDINARY MEMBERS OF COUNCIL

H. SHAW, D Sc.	W S STILES, D Sc., Ph D
E. R. DAVIES, B Sc	C H COLLIE, M A., D Phil
W B. MANN, Ph D.	H. R ROBINSON, D.Sc , Ph D , F.R.S
A J. PHILPOT, O B E., M A , B.Sc	J D BERNAL, M A , F R S
W. D. WRIGHT, D Sc	G I. FINCH, M B.E , D Sc , F R S
B. CHALMERS, D.Sc., Ph.D	D ROAF, M A , D Phil.

EDITORS

Proceedings	Reports on Progress in Physics
J. H. AWBERRY, B.A., B.Sc.	W. B. MANN, Ph.D

COLOUR GROUP

Chairman	Honorary Secretary
R K SCHOFIELD, M.A , Ph.D	W D. WRIGHT, D.Sc

OPTICAL GROUP

Chairman	Honorary Secretary
Instr -Capt. T Y BAKER, B.A R N (ret)	F W H SEWUN R S.

CONTENTS

Part 1. 1 January 1945

	PAGE
Professor E. N. DA C. ANDRADE, Ph D, D Sc, F R S, President of the Physical Society, 1943-45	<i>frontispiece</i>
J. S. PRESTON and G. W. GORDON SMITH The internal resistance of the selenium rectifier photocell, with special reference to the sputtered metal film	1
R. F. BARROW and D. V. CRAWFORD A note on the spectrum of MgO	12
T. VICKERSTAFF The brightness of present-day dyes	15
H. G. HOWELL Resonance in precessional states of diatomic molecules	32
H. G. HOWELL The spectra of tin and lead hydrides	37
R. P. BELL A problem of heat conduction with spherical symmetry	45
S. P. CHOONG Coloration and luminescence produced by radium rays in the different varieties of quartz, and some optic properties of these varieties	49
L. PINCHERLE The polarizing angle for reflection at the boundary between two absorbing media	56
T. J. REHFISCH Rectangular voltage waves from a low impedance source	60
Discussion on papers by L. S. GODDARD (<i>Proc. Phys. Soc.</i> 56, 372 (1944)) and L. S. GODDARD and O. KLEMPERER (<i>ibid.</i> 56, 378 (1944)) on Electron microscopes	63
Corrigendum	66
Recent reports and catalogues	66

Part 2 1 March 1945

W. E. BALLARD The formation of metal-sprayed deposits	67
J. L. HOUGHTON Combinations of spherical lenses to replace non-spherical refracting surfaces in optical systems	84
D. H. SMITH The non-reflecting termination of a transmission line	90
E. C. CRAVEN A study of the comparative method of determining gaseous refractivities	97
A. TAYLOR and H. SINCLAIR The influence of absorption on the shapes and positions of lines in Debye-Scherrer powder photographs	108
A. TAYLOR and H. SINCLAIR On the determination of lattice parameters by the Debye-Scherrer method	126
JOHN JOB MANLEY Recent improvements in a precision balance and the efficacy of rhodium plating for standard weights	136
Reviews of books	145
Corrigendum	146
Recent reports and catalogues	146

Part 3 1 May 1945

R. F. BISHOP, R. HILL and N. F. MOTT The theory of indentation and hardness tests	147
J. B. NELSON and D. P. RILEY An experimental investigation of extrapolation methods in the derivation of accurate unit-cell dimensions of crystals	160
W. E. PERRY Wall- and salt-absorption corrections in radium-content measurements	178

	PAGE
W E. DUNCANSON and C A. COULSON Theoretical shape of the Compton profile for atoms from H to Ne	190
E. H. LINFOOT Achromatized plate-mirror systems	199
WILLIAM HUME-ROTHERY On Gruneisen's equation for thermal expansion	209
H G. W. HARDING Illuminants for colorimetry and the colours of total radiators	222
H MIKHAIL and Y L. YOUSEF. On the measurement of the effective area of a search coil	238
Discussion on paper by W E. BALLARD entitled "The formation of metal-sprayed deposits"	242
Obituary notices	
Sir ARTHUR EDDINGTON	244
CHARLES GLOVER BARKLA	249
JOHN RODERICK ENNIS SMITH	253
MARIUS HANS ERIK TSCHERNING.	253
JOHN KEATS CATTERSON-SMITH	254
Reviews of books	255
Recent reports and catalogues	258

Part 4 1 July 1945

F. W. ASTON, F.R.S., Twenty-first Duddell Medallist	facing 259
C. N. DAVIES. Definitive equations for the fluid resistance of spheres	259
MEGHANAD SAHA, F R S A physical theory of the solar corona	271
T SMITH, F R S On tracing rays through an optical system	286
DAVID OWEN Directional loci in a magnetic field, and the locating of neutral points	294
A. L. M'AULAY and F D. CRUICKSHANK A differential method of adjusting the aberration of a lens system	302
A J. GUINIER Imperfections of crystal lattices as investigated by the study of x-ray diffuse scattering	310
A MORRIS THOMAS and W. L. GENT Permeation and sorption of water vapour in varnish films.	324
F. D. CRUICKSHANK. A system of transfer coefficients for use in the design of lens systems : I. The general theory of the transfer coefficients	350
F. D. CRUICKSHANK. A system of transfer coefficients for use in the design of lens systems : II A second-order correction term	362
Reviews of books	368
Corrigenda	370

Part 5 1 September 1945

K. MENDELSSOHN. The frictionless state of aggregation	371
A. M. BINNIE. A double-refraction method of detecting turbulence in liquids	390
F. GILBERT BROWN. Exact addition formulae for the axial spherical aberration and curvature of field of an optical system of centred spherical surfaces	403
Y. L. YOUSEF. Dynamic measurement of Young's modulus for short wires	412
F. D. CRUICKSHANK A system of transfer coefficients for use in the design of lens systems : III. The contributions to the image aberrations made by the individual surfaces of a lens system	419
F. D. CRUICKSHANK A system of transfer coefficients for use in the design of lens systems : IV. The estimation of the tolerances permissible in the production of an optical system	426

Contents

v

PAGE

F. D. CRUICKSHANK	A system of transfer coefficients for use in the design of lens systems. V. Transfer coefficients for the astigmatism at small aperture and finite obliquity	430
A. L. M'AULAY	A transfer method for deriving the effect on the image formed by an optical system from ray changes produced at a given surface	435
E. F. CALDIN	The relation between the brightness and temperature of a total radiator	440
Reviews of books		444

Part 6 1 November 1945

RAGNAR GRANIT	The electrophysiological analysis of the fundamental problem of colour reception. Fourteenth Thomas Young Oration	447
ARTURO DUPEIER	The geophysical aspect of cosmic rays. Twenty-ninth Guthrie Lecture	464
J. B. NELSON and D. P. RILEY	The thermal expansion of graphite from 15° c. to 800° c. Part I. Experimental	477
D. P. RILEY	The thermal expansion of graphite: Part II. Theoretical	486
A. CHARLESBY	Structure and orientation in thin films of polythene	496
A. CHARLESBY	Effect of temperature on the structure of highly polymerized hydrocarbons	510
J. C. JAEGER	Note on diffusion in the ionosphere	519
S. R. PELC	The photographic action of x rays	523
D. H. SMITH	A method for obtaining small mechanical vibrations of known amplitude	534
T. SMITH, F.R.S.	Theoretical investigation on telephoto lenses	543
T. SMITH, F.R.S.	Variational formulae in optics	558
F. W. CUCKOW	The electron microscope and its applications	564
C. R. BURCH, F.R.S.	Flat-fielded singlet aplanats	567
Demonstrations.		
A. Franck and Hertz	critical potential experiment, by J. H. SANDERS	577
Apparatus illustrating the Michelson stellar interferometer, by P. J. TREANOR, S.J.		579
Obituary notices		
JOHN AMBROSE FLEMING		581
SIDNEY SKINNER		584
LIONEL ROBERT WILBERFORCE		585
JAMES YOUNG		585
Reviews of books		586
Index to Volume 57		591
Index to Reviews of books		596
Proceedings at the Meetings, Session 1944-45		vi
Report of Council for the year 1944		xiii
Report of the Honorary Treasurer for the year 1944		xvii
E. N. da C. ANDRADE, F.R.S.	The history and future of the Physical Society. Presidential Address	xxi

PROCEEDINGS AT THE MEETINGS OF THE PHYSICAL SOCIETY

SESSION 1944-45

20 September 1944

The eighteenth meeting of THE COLOUR GROUP, at the Lighting Service Bureau, Savoy Hill, London W.C. 2, Mr J. Guild being in the Chair.

A lecture entitled "Illuminants for colorimetry and the colours of total radiators" as delivered by H. G. W. Harding, B.Sc., and was followed by an informal discussion.

29 September 1944

The thirteenth meeting of THE OPTICAL GROUP, at Imperial College, London S.W. 7, Professor A. O. Rankine being in the Chair

A demonstration of "High-intensity projectors for stereoscopic screen projection" was given by A. C. W. Aldis, M.A.

The following papers were read and discussed

"The monochromatic aberrations", by H. H. Hopkins, Ph.D. ;

"Transfer method for deriving the effect on the image formed by an optical system from ray changes produced at a given surface", by Professor A. L. M'Aulay

"A differential method of adjusting the aberration of a lens system", by Professor A. L. M'Aulay and F. D. Cruickshank, B.Sc.

"A system of transfer coefficients for use in the design of lens systems", by F. D. Cruickshank, B.Sc.

(A summary of the last three papers was presented by Dr C. G. Wynne.)

24 October 1944

Science Meeting, at the Cavendish Laboratory, Cambridge, the President, Professor E. N. da C. Andrade, being in the Chair

A lecture on "Active nitrogen" was delivered by Professor S. K. Mitra, D.Sc., and was followed by an informal discussion

27 October 1944

Science Meeting, at Imperial College, London S.W. 7, the President, Professor E. N. da C. Andrade, being in the Chair

The following were elected to Fellowship: John Norman Aldington, Arthur Joseph Hughes, Cyril James Henry Monk, Wilfred Ernest Watson-Baker, Charles Gorrie Wynne.

It was announced that the Council had elected the following to Student Membership: Rafael Armenteros, Gordon John Bell, Maurice Arthur Cayless, Rex Herbert Coffin, Montague Cohen, Derek H. Cole, Dennis Henry Colvill, Robert William Cranston, Peter Crosby, Michael William Feast, Lily Irene Gadd, Malcolm Hobson Green, Joan Irene Harper, Ronald Bertie Ide, Clifford Philip Newman, Donald Perkins, Thomas Henry Richardson, Peter Henry Rose, Hilmi Samara, Vera Mary Scamans, Beta May Sparkes, John Henry Strong, George Geoffrey Twidle, R. J. Wakelin

Demonstrations of "Rectangular voltage waves from a low-impedance source" were given by T. J. Rehfish.

The following papers were read and discussed :

"The computation of electron trajectories in axially symmetric fields", by L. S. Goddard, B Sc ;

"Electron ray tracing through magnetic lenses", by L. S. Goddard, B.Sc, and O Klemperer, Dr Phil.

31 October 1944

Science Meeting, held jointly with the British Institution of Radio Engineers in the rooms of the Institution of Structural Engineers, Upper Belgrave Street, London S W 1, Mr L. McMichael being in the Chair.

A lecture entitled "Physics and radio" was delivered by Professor E N. da C Andrade, F R S, and was followed by an informal discussion.

9 November 1944

Science Meeting, at The Royal Society, Burlington House, London W. 1, the President, Professor E N. da C Andrade, being in the Chair.

The following were elected to Fellowship : William George Atkins, Reginald John Bartle, Cyril Arnold Bridgewater, Cyril Bristow, Kenneth George Brummage, James Alexander Burns, Clifford Charles Butler, Michael Anthony Grace, Geoffrey Grime, R. W James, Sidney Gabriel Richard Lindboe, Kenneth Alexander McMurtrie, Jozef Mazur, Mohamed Gamal El-Din Nooh, Hubert Derrick Parbrook, Jerzy Skrebowski, Robert Harbinson Sloane, Andrew David Thomas, Charles Edney Wheeler, Walther Wolff

The President delivered the following address

"We welcome to-day certain of our French comrades in physics, members of our sister society, la Société française de Physique. We have here Professor G. A. Boutry, Professor P. Auger and Dr. S. Rosenblum. Through the kind offices of M Boutry, professeur au Conservatoire national des Arts et Métiers, and a Fellow of our Society, I have received a brief account of the history of the Société since the outbreak of war. In it the President, Professor J. Cabannes, records the lamentable fate of many members whose names are known to, and honoured by, all of us. I will ask you to stand, as a token of respect, while I read the names of those who have died

G. Dechêne, J. Farineau, H. du Mesnil du Buisson and J. Rossignol fell on the field of honour in the early battles.

F. Holweck, who made his name famous by his pump and by other ingenious instruments, by his work on soft x rays and by other notable researches, was murdered by the Gestapo ; and J. Solomon, one of the most promising of the younger theoretical physicists, was shot by the German military authorities.

Deaths that we have to deplore, some no doubt accelerated by the traditional brutalities of the Germans, are those of H. Buisson, H. Chipart, A. Dufour, A. Guillet, Victor Henri, Jean Perrin (Foreign Member of the Royal Society and Honorary Fellow of our Society), Emile Picard (likewise a Foreign Member of the Royal Society) and Pierre Weiss.

Besides those who are known definitely to have lost their lives, there are others whose ultimate fate is unknown. H Abraham, Eugene Bloch, G. Bruhat, L. Cartan, C. Sadron and J. Yvon have been deported to Germany, and we dare scarcely hope that they have found humane treatment there. Of Paul Langevin, who received in 1940 the highest honour that the Royal Society has to give, the Copley Medal, we have no news, unless M. Boutry has something to tell us

Let me assure our French friends that although, as Mr. Eden said yesterday, 'People in Britain do not yet fully understand how complete, how merciless, how dastardly has been the devastation inflicted by the German armies in Allied lands

as they withdrew', some of us here have some comprehension of German brutality and are not anxious to welcome back in our midst the representatives of German science who also represent German unscrupulousness, German opportunism and German dishonour.

I am glad to be able to inform the Physical Society that all through the troublous days of the German occupation, zealous and courageous French colleagues have kept alive the science of physics in France. In June 1941 appeared a new publication, the *Cahiers de Physique*, and the first number of another new publication, the *Annales de Géophysique*, is at this moment in the press. We congratulate our French brethren on the success of their strenuous efforts.

M. Boutry, Au nom de la Physical Society je vous prie de bien vouloir saluer de notre part nos camarades de la Société française de Physique et de leur dire combien nous regrettons le lamentable sort de nos confrères, victimes du maudit système Nazi, ou plutôt allemand, combien nous espérons voir renaître dans toute leur gloire traditionnelle et la France et la physique française. Nous avons remarqué avec la plus vive émotion comment, au milieu de tant de dangers et de difficultés, les physiciens français ont continué leurs travaux désintéressés. Tout ce que la Physical Society peut faire pour encourager et appuyer nos collègues, nos amis français, soyez assuré que ce sera fait. C'est pour nous un très grand honneur de pouvoir désormais renouveler nos cordiales et traditionnelles relations avec la Société française de Physique. Cher collègue, soyez-le bienvenu."

A lecture on "The night sky" was delivered by Professor S. K. Mitra, D.Sc., and was followed by an informal discussion.

22 November 1944

The nineteenth meeting of THE COLOUR GROUP, at Imperial College, London S.W. 7, Mr. J. Guild being in the Chair.

A lecture entitled "Retinal structure and colour vision" was delivered by E. N. Willmer, M.A., B.Sc., and was followed by an informal discussion.

23 November 1944

Science Meeting, at the Royal Society, Burlington House, London W. 1, the President, Professor E. N. da C. Andrade, being in the Chair.

Kenneth Conrad Bryant, Deryck Chesterman and Peter Savic were elected to Fellowship.

The President announced the death, on the previous day, of Sir Arthur Eddington, O.M., F.R.S., (President 1930-32) and read an account of Sir Arthur's life and work.

A lecture on "A physical theory of the solar corona" was delivered by Professor M. N. Saha, D.Sc., F.R.S., and was followed by an informal discussion.

24 November 1944

The fourteenth meeting of THE OPTICAL GROUP, at Imperial College, London S.W. 7, Professor A. O. Rankine, being in the Chair.

The following papers were read and discussed.

"Notes on the evolution of the inverting eyepiece", by E. Wilfred Taylor,

"Achromatism of two thin separated lenses and of a cemented doublet", by Instr.-Capt. T. Y. Baker.

A demonstration of "A new modification of a ray plotter" was given by Mr. B. K. Johnson.

19 January 1945

Science Meeting, at The Royal Society, Burlington House, London W 1, the President, Professor E. N. da C. Andrade, being in the Chair.

The following were elected to Fellowship. William Fordham Cadmore Cooper, Zygmunt Alexander Klemensiewicz, Kenneth Herbert Spring, Harry Verdon Stopes-Roe, William John Wilson.

It was announced that the Council had elected the following to Student Membership. Denys Osmund Akhurst, Charles Anthony Arkell, Gordon Burnand, Alfred Norman Burton, David Keith Butt, Derek Young Coomber, Christian Ellis Coulman, Ronald Dee, Gerald Morley Habberjam, Roy Hayes, Gerald Alfred Jackson, John Lister, Harold Frederick William Mander, Alan Pearce Morgan, Clifford Philip Newman, Arthur Coates Parker, Thomas Frederick Porter, John Alexander Pryde, Geoffrey Cecil Pyle, Derek William Saunders, Geoffrey George Smith, Derek Malcolm Thomas, Maurice William Tulett, Sidney John Walton, John Woods, Peter Robert Wyke

The President read an appeal, copies of which were afterwards posted to Fellows and friends of the Society, for contributions to a Holweck Prize Fund

A lecture on "Imperfections of crystal lattices as investigated by study of x-ray diffuse scattering" was delivered by Dr A. J. Guinier and was followed by an informal discussion.

7 February 1945

The twentieth meeting of THE COLOUR GROUP, at the School of Photo-Engraving and Lithography, Bolt Court, Fleet Street, London E C 4, Mr J. Guild being in the Chair.

Mr. H. M. Cartwright gave a lecture and demonstrations on "Colour printing and problems of colour reproduction", which were followed by an informal discussion

16 February 1945

The fifteenth meeting of THE OPTICAL GROUP, at Imperial College, London S W. 7, Professor A. O. Rankine being in the Chair.

The following papers were read and discussed

"The properties of some new optical glasses", by R. E. Bastick, Ph D.

"Routine refractometry for optical glasses", by W. C. Hynd, M.Sc.

23 February 1945

Science Meeting, at the Science Museum, London S W. 7, the President, Professor E. N. da C. Andrade, being in the Chair

The following were elected to Fellowship, all but the first ten being transferred from Student Membership. Ernest Reginald Adams, Reginald Joseph Cole, Ernest Cecil Craven, Irena Gimpel, Franklin Stewart Harris, Charles Robert Sinden Manders, Ronald Gladstone Mitton, Leonard Eugene Newnham, Albert Hickford Turnbull, Edward Leighton Yates; Alan Wawn Agar, Norman Rushton Bailey, Roderick Kirkwood Barnes, Dennis Stanley Beard, Clifford Henry James Beaven, Margaret Elizabeth Jane Carr, Isabel Helen Cox, Donald Watts Davies, Joseph William Fox, Keith Davy Froome, Richard Mead Goody, Norman John Harris, Kenneth William Hillier, Mervyn Guy Newcomen Hine, George Francis Hodsman, Francis Raymond Holt, Denys Sydney Hopper, Howard Arthur Hughes, Robert William Gainer Hunt, William Alan Jennings, Peter Reginald Layton, Ian Alexander Darroch Lewis, Derrick Joseph Littler, Emlyn Howard Lloyd, Thomas Alan Minns, Peter Lloyd Morgan, John Pollard, William Leslie Roberts, James Richard Stansfield, Denis Warwick Stops, Bernard Sugarman, Rudolf Arnold Wiersma.

Proceedings at meetings

It was announced that the Council had elected the following to Student Membership : Fritz K Bauchwitz, Michael Somerset Ridout, Keith Varnden Roberts, John Roy Schofield, Derek Anthony Silverston, Wladyslaw Jerzy Swiateki

A lecture on "The electron microscope and its applications" was delivered by F W. Cuckow, B Sc., and was followed by an informal discussion.

9 March 1945

Science Meeting, at Imperial College, London S W 7, Brigadier J L P MacNair, Fellow, being in the Chair

The following were elected to Fellowship, the last four being transferred from Student Membership : Bengt Edlén, Adolf Schallamach, Claude Ivan Snow, William Ernest Arnould Taylor ; Harold Raynor Allan, Alan Stephen Edmondson, Roy Heastie, Geoffrey Sherwood Waters

The following papers were read and discussed

"The formation of metal-sprayed deposits" (with demonstration), by W E Ballard, F.R.I.C ;

"The theory of indentation and hardness tests", by R F Bishop, B A , R Hill, B Sc., and N F. Mott, M.A , F R S

21 March 1945

The twenty-first meeting of THE COLOUR GROUP, at Imperial College, London S W. 7, Mr. J. Guild being in the Chair.

A paper on "The presentation of the C I E. system of colour specification" was read by R K Schofield, M A., Ph.D , and was followed by an informal discussion

The meeting was preceded by the fifth *Annual General Meeting of the Colour Group* for the presentation of the Committee's report on the work of the Group in 1944-45 and for the election of Officers and Committee for 1945-46

20 April 1945

The sixteenth meeting of THE OPTICAL GROUP, at Imperial College, London S W 7, Instr.-Capt. T Y. Baker being in the Chair

The following papers were read and discussed

"A theoretical investigation on telephoto lenses", by T Smith, M A , F R S ,

"On tracing rays through an optical system" (fifth paper), by T Smith, M.A , F R S ;

"Exact addition formulae for the axial spherical aberration and curvature of field of an optical system with centred spherical surfaces", by F Gilbert Brown.

The meeting was preceded by the fourth *Annual General Meeting of the Optical Group* for the presentation of the Committee's report on the work of the Group in 1944-45 and for the election of Officers and Committee for 1945-46

23 May 1945

Science Meeting, at the Royal Institution, Albermarle Street, London W 1, the President, Professor E. N. da C. Andrade, being in the Chair.

The following were elected to Fellowship, all but the first three being transferred from Student Membership . Raymond Frederick Cyster, Hubert Leslie Hulland, Archibald Charles Lock , Arthur Reginald Bevan, Jack Dunkley, Clifford McDonald Hargreaves, George Albert Mann, Ernest Henry Putley, Alan Stott, Laurence William Stevens Wilson.

It was announced that the Council had elected the following to Student Membership : James Howard Barrett, Cyril Eugene Challice, Francis Maurice Comerford, Peter Brian Curtis, Reginald Leslie Gleave Gilbert, Rita Joyce Hancock, Hugh Donald Hemmer, Ralph Ewart Hickman, James Alexander Leys, Peter Murden, Raymond Peter William Scott

The twenty-first Duddell Medal was presented to Dr F. W. Aston, F.R.S., in recognition of his invention and development of the mass spectrograph.

Professor Andrade delivered his Presidential Address on "The history and future of the Physical Society"*

* This volume, p. xvi

23 May 1945

Annual General Meeting, at the Royal Institution, Albermarle Street, London W. 1, the President, Professor E. N. da C. Andrade, being in the Chair.

The minutes of the previous Annual General Meeting and of the Extraordinary General Meeting, held on 24 May 1944, were read and accepted as correct

The reports of the Council and the Honorary Treasurer and the Annual Accounts for 1944 were adopted.

The Officers and Council and the Auditors for 1945-46 were elected

Votes of thanks were accorded to the Rector and Governing Body of Imperial College, the President and Council of the Royal Society, the Managers of the Royal Institution, and the Director of the Science Museum for excellent accommodation at meetings; to the Royal Commission for the Exhibition of 1851 and Dr Evelyn Shaw for the office and library accommodation at 1 Lowther Gardens, Exhibition Road, London S.W. 7; to Dr. A. F. C. Pollard for preparing the U D C Index Slips for the *Proceedings*, and to the retiring Officers and Council

30 May 1945

The twenty-second meeting of THE COLOUR GROUP, at the Royal Photographic Society, 16 Prince's Gate, London S W 7, Dr. R. K. Schofield being in the Chair.

A lecture on "Colour television" was delivered by L. C. Jesty, B.Sc., M.I.E.E., and was followed by an informal discussion

8 June 1945

Science Meeting, at Imperial College, London S.W. 7, the President, Professor D. Brunt, being in the Chair.

The following were elected to Fellowship, the last four being transferred from Student Membership : Ronald George Allen, Jack Dainty, Douglas William Nicol Dolbear, André Jean Guinier, Harold Horace Hopkins, Richard Albert Hull, Günther Loeck, Hans Motz, Stefan Richard Pelc, Harry Hemley Plaskett, Robert Arthur Scott, David Starkie, Leonard Walden, Frederick Henry East, Arnold Stanley Knight, Douglas Henry Pearson, Harry Pratt

The President read a letter from Professor C. Fabry giving further bad news of the fate of French physicists, supplementing that given at the meeting of 9 November 1944.

The following demonstrations were given and discussed

"Directional loci in a magnetic field and the locating of neutral points", by D. Owen, B.A., D.Sc.,

"A simple form of refractometer of the Jamin type", by H. G. Kuhn, Dr.Phil., and G. A. Wheatley, B.A.;

"A demonstration model of Michelson's stellar interferometer", by P. J. Treanor, S.J., B.A.;

"A Franck-and-Hertz critical-potential experiment, arranged for demonstration with a cathode-ray oscillograph", by J. H. Sanders, B.A.,

"A water-operated model klystron", by W. J. Scott

22 June 1945

The seventeenth meeting of THE OPTICAL GROUP, at Imperial College, London S.W. 7, Instr.-Captain T. Y. Baker being in the Chair

A lecture on "Instrument design" was delivered by A. F. C. Pollard, D. Sc

29 June 1945

Science Meeting, at the Royal Institution, Albemarle Street, London W. 1, the President, Professor D. Brunt, being in the Chair.

The following were elected to Fellowship, the last two being transferred from Student Membership: John Frank Allen, Samuel Thomas Adrian Hall, George Arkla Haile, Oscar Martin Lee, Thomas Henry Pearn, Leslie Ward, George Matthew Leak, Leslie Hamilton Tarrant

It was announced that the Council had elected the following to Student Membership: Norman Brooke, Donald H. Brown, Eric Lionel Ferguson, Sylvia Mary Gumbrell, Ronald Mayoh, William Stanford Michael.

The President announced that the Foreign Secretary, Professor E. N. da C. Andrade, was representing the Society at the meetings in Moscow and Leningrad to celebrate the 220th anniversary of the foundation of the Academy of Sciences of the U.S.S.R.

The fourteenth Thomas Young Oration was delivered by Professor Ragnar Granit, who took as his subject: "The electrophysiological analysis of the fundamental problem of colour reception".

6 July 1945

Science Meeting, at the Royal Institution, Albemarle Street, London W. 1, the President, Professor D. Brunt, being in the Chair

The following were elected to Fellowship: William Cyril Barry, Owe Berg, Herbert Cornes, Thomas Peter Taylor, Henry William H. Warren

The twenty-ninth Guthrie Lecture was delivered by Professor Arturo Duperier, who took as his subject: "The geophysical aspect of cosmic rays".

REPORT OF COUNCIL FOR THE YEAR ENDED 31 DECEMBER 1944

INTRODUCTORY AND GENERAL

The year under review was one of steady progress in difficult circumstances and of preparation for much greater activities in the not far distant future. The membership has increased, the Science Meetings have been remarkably successful, and the Colour and Optical Groups have continued their very useful work without setback or slackening of pace.

It has again proved possible to maintain a satisfactory financial position without increasing the basic rates of Fellows' and Student Members' annual subscriptions or the rate of outside subscriptions to the *Proceedings*.

Preliminary steps have been taken towards the resumption of the Society's Annual Exhibition of Scientific Instruments and Apparatus, the 30th exhibition being planned for January 1946.

Much consideration has been given during the year to the co-ordination of the activities of our Society and the Institute of Physics, especially in regard to post-war publications and the formation of Groups for the discussion of certain subjects. Steps have been taken to set up a Joint Committee of the two bodies to discuss these and other matters of common interest and importance.

The Society has supported the work of the Parliamentary and Scientific Committee by taking up corporate membership, and is represented on the Executive Committee by Professor E. N. da C. Andrade, who has been elected a Vice-President of the Committee.

The Council has decided that an application for the incorporation of the Society under Royal Charter shall be made when such a course becomes possible.

Preliminary consideration has been given to the question of the future housing of the Physical and other scientific Societies. The President, Professor E. N. da C. Andrade, represented the Society on a delegation, headed by the President of the Royal Society, which was received on 13 October 1944 by the Lord President of the Council, the Chancellor of the Exchequer and the Minister of Works and Buildings. Statements of the needs of the societies concerned are now being prepared and considered.

Through the generosity of the Royal Commission for the Exhibition of 1851, the Society continues to have excellent office and library accommodation at 1 Lowther Gardens, South Kensington. Acute shortage of office staff hampered the day-to-day work of the Society for nearly the whole year, but such difficulties will, it is hoped, be resolved in the near future.

MEETINGS

An Annual General Meeting was held at Imperial College on 24 May 1944 for the presentation of the 1943 Reports of the Council and the Honorary Treasurer, and for the election of the Officers and Council for 1944-45. This was followed by an Extraordinary General Meeting for the election of Professor A. F. Joffé, Vice-President of the Academy of Sciences, Moscow, as an Honorary Fellow of the Society.

Eleven Science Meetings were held during the year: four at Imperial College, two in the rooms of the Royal Society, one each at the Royal Institution, the Clarendon Laboratory and the Cavendish Laboratory, and two (jointly with other bodies) elsewhere. Two of the meetings were devoted to the Guthrie Lecture noted below, and three to special lectures on "Active Nitrogen" and "The Night Sky" by Professor S. K. Mitra, and "A Physical Theory of the Solar Corona" by Professor M. N. Saha (*Proceedings*, 57 (1945), p. 271). At three meetings discussions were held on "Band Spectra and Energies of Dissociation of Diatomic Molecules", opened by Dr. A. G. Gaydon (*Proceedings*, 56 (1944), p. 204), and "Physics and the Scientific Instrument Industry", opened by Mr. A. J. Philpot (*Proceedings*, 56 (1944), pp. 263 and 396), and a lecture-survey on "Current Problems of Visual Research" was delivered by Dr. W. S. Stiles (*Proceedings*, 56 (1944), p. 329). At one meeting Demonstrations of Rectangular Voltage Waves were given by Dr. T. J. Rehfisch (*Proceedings*, 57 (1945), p. 60) and papers on Electron Trajectories and Ray-tracing by Mr. L. S. Goddard and Dr. O. Klemperer were read and discussed (*Proceedings*, 56 (1944), pp. 372 and 378, and 57 (1945), p. 63). One meeting was held

jointly with the British Institution of Radio Engineers for a lecture by Professor E. N. da C. Andrade entitled "Physics and Radio", and one jointly with the Plastics Group of the Society of Chemical Industry for a lecture by Dr. W. T. Astbury on "X-ray Examination of Plastics".

The liberation of France has been quickly followed by the renewal of contact with our colleagues of the Société française de Physique. This was made the occasion of a brief but moving ceremony at the Science Meeting on 23 November*.

The Council records its thanks for hospitality extended during the year to the Society and its two Groups by the President and Council of the Royal Society, by the Rector and Governing Body of Imperial College, by the Managers of the Royal Institution, by the Electric Lamp Manufacturers' Association and by the Director of the Science Museum.

GUTHRIE LECTURE

The twenty-eighth Guthrie Lecture was delivered on 26 April at the Royal Institution and on 29 April at the new Clarendon Laboratory, Oxford, by Professor Joel H. Hildebrand (University of California), who took as his subject "The Liquid State" (*Proceedings*, 56 (1943), p. 221).

RUTHERFORD MEMORIAL LECTURE

The second Rutherford Memorial Lecture, on Rutherford's work after the Manchester period, should have been delivered by Professor J. D. Cockcroft in the early autumn of 1944, but was unavoidably postponed on account of the Lecturer's prolonged absence from this country. Arrangements will be made for the Lecture as soon as possible.

DUDELL MEDAL

The Council awarded the twenty-first Duddell Medal to Dr. F. W. Aston in recognition of his invention and subsequent improvement of the mass spectrograph. The presentation of the Medal could not be made in 1944, but will take place at a Science Meeting as early as can be arranged.

THE R W PAUL INSTRUMENT FUND

In accordance with the will of the late Mr Robert W. Paul, a Fund, consisting of Ordinary Shares in the Cambridge Instrument Company valued for probate at £110,036 10s., has been established for the provision of facilities (equipment, building, staff or financial aid) for "the design, construction and maintenance of novel, unusual or much improved types of physical instruments and apparatus needed for an investigation in pure and applied physical science and . . . the advance of knowledge and the arts, where a relatively large expenditure may be justifiably risked on experimental apparatus . . ." The application of the income is to be directed, controlled and administered by a Committee of six members, of whom three are appointed by the Royal Society and one each by the Physical Society, the Institution of Electrical Engineers and the Institute of Physics. If no grant is made for a period of seven consecutive years, the capital, future income and accumulation of income are thereafter to be held in trust for the Royal Society, the Royal Institution and the Physical Society in equal shares for their general purposes.

The Council has appointed Professor A. O. Rankine as the representative of the Physical Society on the Committee.

CHARLES VERNON BOYS PRIZE

By the will of the late Sir Charles Boys, a former President, a nett legacy of £900 was bequeathed to the Society, and the Council has decided that the interest shall be used for the institution of the Charles Vernon Boys Prize, to be awarded annually, without restriction of nationality or of Fellowship of this or any other Society, for distinguished research in experimental physics, either still in progress or carried out within the five years preceding the date of the award. The first award of the Prize will be made in 1945.

* The address delivered by the President on this occasion is recorded under "Proceedings at the Meetings" on pp vii-viii of this volume.

ADDENBROOKE BEQUEST

Through the generosity of Mrs. B. E. F. Addenbrooke, and by the wish of her late husband, Mr. G. L. Addenbrooke, a Fellow of the Society, Government Stock to the value of about £315 has been bequeathed to the Society. With the approval of the executors, the Council has decided that the interest on this bequest shall be used in the first instance for an annual purchase, starting in 1945, of standard books and other books of outstanding merit and usefulness on "the structure, properties and behaviour of non-metallic substances". The volumes so purchased will be suitably inscribed and placed in a special section of the Society's Library, and their titles will be announced in the Notices to Members and communicated to appropriate bodies.

PROCEEDINGS

Volume 56 (1944) of the *Proceedings* appeared in the usual six two-monthly parts, but was about 20 per cent smaller than the 1943 volume; the reduction in size has made possible a desirable increase in the number of copies printed, although the official allowance of paper has remained unchanged. The increased charges for advertisements reported a year ago came into effect by the end of the year; the space taken was exactly the same as in the 1943 volume. Two more Parts of Volume 54 (1942) which had run out of stock have been reprinted.

The Council is again deeply indebted to Professor A. F. C. Pollard for his invaluable help in the preparation of the Universal Decimal Classification Index Slips, which are supplied with the *Proceedings* to all members and subscribers who make application for them in advance.

REPORTS ON PROGRESS IN PHYSICS

While physicists remain so busily engaged on work of the highest importance for the Allied cause, and while the supply of paper remains so meagre, it has again been decided that the next Volume must, like its immediate predecessor, cover two years instead of one; volume 10 (1944-45) is in active preparation and should be ready early in 1946. The sales of all the available volumes, 5 (1938) - 9 (1942-43), have continued to be very satisfactory, and the remaining stock of Vol. 5 became exhausted towards the end of the year. Vol. 4 (1937), which has been out of print for several years, but which is still in big demand, is being reprinted, and should be available in the summer of 1945. The Council again records its thanks to Miss M. M. Mitchell, the Publications Manager of the American Institute of Physics, for her valuable co-operation in the supplying of copies of the *Reports* to American physicists.

REPRESENTATION ON OTHER BODIES

The representatives of the Society on the National Committees for Physics and for Scientific Radio, on the Committee of Management of *Science Abstracts*, on the Board of the Institute of Physics and on the Council of the British Society for International Bibliography were the same in 1944 as were reported a year ago. As mentioned elsewhere in this Report, the Society is represented by Professor E. N. da C. Andrade on the Parliamentary and Scientific Committee and by Professor A. O. Rankine on the R. W. Paul Instrument Fund Committee.

MEMBERSHIP

A limited edition of the List of Officers and Members, corrected to 30 September 1944, has been printed. It is satisfactory to be able to record a notable growth of membership for the fourth year in succession. As the following table shows, the Fellowship increased in 1944 by 5.3 per cent and the Student Membership by nearly 3 per cent, making an increase in the total membership of nearly 5 per cent, which is less than in 1942 and in 1943. The most welcome feature of the table is the large number of newly elected Fellows, which was exceeded in only one previous year, 1942.

Roll of Membership		Hon. Fellows	Hon. Fellows, Optical Society	Ex-officio Fellows	Fellows	Student Members	Total
Totals, 31 Dec. 1943		10	3	4	1032	271	1320
Changes during 1944	Newly elected	1			56	23	52
	Transferred				23	1	
	Deceased		1		12	3	44
	Resigned				7	17	
	Lapsed				4		
	Suspended				1		
	Net increase	1	-1		55	8	63
Totals, 31 Dec. 1944		11	3	4	1087	279	1383

OBITUARY

The Council records with regret the deaths of the following Fellows :—Mr. C. T. Archer, Sir Charles V. Boys, (Past President) Mr. J. W. Bullerwell, Mr. A. W. Claydon, Sir Arthur S. Eddington (Past President), Mr. J. C. R. Le Manuais, Sir Thomas R. Lyle, Professor H. F. Newall, Dr. J. K. Roberts, Mr. A. C. Shearman, Principal S. Skinner, Professor M. L. E. Tscherning (Honorary Fellow of the Optical Society), Professor L. R. Wilberforce, and the death of a Student Member, Dr. J. R. Ennis Smith.

COLOUR GROUP

At the fourth Annual General Meeting of the Group, which was held at Imperial College on 8 March 1944, Mr. J. Guild was re-elected as Chairman, Dr. W. D. Wright was elected as Honorary Secretary, and the Committee for 1944-45 was elected.

Five Science Meetings were held during the year. Three of them were at Imperial College on 26 January, 8 March and 22 November, the subjects discussed being "The brightness of present-day dyes" (paper by Dr. T. Vickerstaff, *Proceedings*, 57 (1945), 15), "Colour vision deficiency in industry" (paper by Professor J. H. Sharby), and "Retinal structure and colour vision" (paper by Mr. E. N. Willmer). A meeting was held at the Royal Society of Arts on 7 June for a paper on "The Munsell system" by Dr. W. D. Wright, and one at the Lighting Service Bureau of the Electric Lamp Manufacturers' Association for a lecture on "Illuminants for colorimetry and the colours of total radiators" by Mr. H. G. W. Harding (*Proceedings*, 57 (1945), 222).

The reports of the Sub-Committees on Colour Terminology and Colour Vision Deficiency in Industry are nearing completion. Close contact with the Inter-Society Colour Council in the U.S.A. has been maintained.

The membership of the Group on 31 December was as follows :

Members of the Physical Society	79
Members of other participating bodies	59
Members of five firms subscribing for sustaining membership	15
Other members	5
Total	158

OPTICAL GROUP

The Third Annual General Meeting of the Group was held at Imperial College on 21 April 1944, when Professor A. O. Rankine and Professor L. C. Martin were re-elected as Chairman and Honorary Secretary, respectively, and the Committee for 1944-45 was elected.

Five Science Meetings were held during the year. At the Northampton Polytechnic on 18 February, Mr. H. H. Emsley delivered a lecture entitled "Some notes on space perception" (*Proceedings*, 56 (1944), 293), and papers on "Contact lenses" by Mr. K.

Clifford Hall and Mr. E. F. Fincham were read and discussed (*Proceedings*, 56 (1944) 272). At a special meeting held jointly with the Scientific and Technical Group of the Royal Photographic Society on 30 May, Mr. E. R. Davies presented for discussion an address entitled "Psychophysics and photography" by Dr. Lloyd A. Jones (*J. Opt. Soc. Amer.* 54 (1944), 66). Three meetings were held at Imperial College on 21 April, 29 September and 24 November; the first was devoted to a lecture entitled "My fifty years in the optical industry" by Mr. J. W. Hasselkus; and the subjects of the other two were demonstrations of "High-intensity stereoscopic screen projection" by Mr. A. C. W. Aldis and "A new modification of a ray plotter" by Mr. B. K. Johnson, papers by Mr. H. H. Hopkins, Instr.-Capt. T. Y. Baker, and Mr. E. Wilfred Taylor ("The inverting eye-piece and its evolution", *J. Sci. Inst.* 22 (1945), 43), and a digest by Mr. C. G. Wynne of papers by Professor A. L. M'Aulay and Mr. F. C. Cruickshank on the design of lens systems (*Proceedings*, 57 (1945), 302, 350, 362).

The membership increased by 10 during the year, and was as follows on 31 December:

Members of the Physical Society	152
Members of other participating bodies	89
Members of twenty-three firms subscribing for sustaining membership	97
Other members	8
Total	346

REPORT OF THE HONORARY TREASURER FOR THE YEAR ENDED 31 DECEMBER 1944

In 1944 the excess of income over expenditure (£1144) was larger than in any previous year, and £566 higher than in 1943; the total expenditure (£4207) was £278 less than a year ago, and the total income (£5351) was £288 more.

The only item of expenditure showing an increase since 1943 was in respect of the Society's occupation of rooms at 1 Lowther Gardens, where it became necessary to engage a fire-watcher for the greater part of the year. By arrangement with the Secretary of the Royal Commission for the Exhibition of 1851, a senior student at Imperial College was employed in this capacity, the expense being shared equally by the Royal Commission and the Society.

Expenditure on secretarial and clerical services was low on account of shortage of staff, which will, it is hoped, be rectified in the near future.

The income from sales of publications was about the same as in 1943; a considerable increase of sales of the *Proceedings* balanced the falls of sales of *Reports on Progress in Physics* and *Special Publications*. Every other important item of income showed an increase. The charges for advertisements in the *Proceedings* were appropriately raised during the year.

The total value of the Society's investments at the end of the year was £2093 higher than at the end of 1943; of this, the newly instituted Charles Vernon Boys Prize Fund accounts for £900, and £1000 was invested in 3% Defence Bonds early in 1944. The market values showed a slight appreciation.

A further, and larger, grant by the Royal Society from the Rockefeller Gift, to assist the Society to meet war-time publication difficulties, is gratefully acknowledged; the grant has been expended on the reprinting of Parts of the 1942 *Proceedings*.

Owing to the stringency of the paper control and the shortage of original material which it is permissible to publish during the war, both the size of the *Proceedings* and also the frequency of publication of the *Reports on Progress in Physics* have been reduced considerably in the last few years. When such limitations are removed the Society will have reason to be pleased with the favourable balance it has accumulated. In the near future it will be necessary also to acquire a considerable amount of office furniture, not only to replace that now in use, which is the property of the Institute of Physics, but also to meet the needs of the increasing activities of the Society.

(Signed) CLIFFORD C. PATERSON,
Honorary Treasurer.

19 April 1945.

INCOME AND EXPENDITURE ACCOUNT FOR THE YEAR ENDED 31 DECEMBER 1944

1943 £	EXPENDITURE	£ s. d.
	T _o	£ s. d.
266	" Science Abstracts "	225 6 6
1687	Normal Publications
67	" Proceedings "	1685 19 11
1119	Agenda Papers and Notices	58 18 0
	" Reports on Progress in Physics "	779 3 9
		<hr/> 2524 1 8
267	Postage on Publications and General Correspondence	234 6 1
53	Expenses at Meetings	61 18 9
40	Honoraria to Special Lecturers	26 5 0
	Administration Expenses	
934	Secretarial and Clerical Assistance and Office Expenses	910 0 10
52	1 Lowther Gardens (including Air Raid Precautions)	224 13 2
		<hr/> 1134 14 0
578	Balance, being excess of Income over Expenditure, carried forward to Accumulated Fund	1144 1 2
		<hr/>
		£5350 13 2

1943 £	INCOME	£ s. d.
	By	£ s. d.
39	Subscriptions:	
	Entrance Fees	54 12 0
1920	Fellows	2004 19 0
140	Student Members	147 10 0
25	Colour Group	22 18 0
58	Optical Group	66 2 0
	For " Science Abstracts " and Advance Proofs	147 14 7
	Sales	<hr/> 2443 15 7
884	" Proceedings "	1019 2 5
903	" Reports on Progress in Physics "	852 3 10
134	Special Publications	47 15 10
		<hr/> 1919 2 1
307	Advertisements in " Proceedings "	333 13 6
283	Dividends from Investments	304 2 0
	Transfer from Life Composition Fund of amounts paid by Fellows now deceased	50 0 0
30	Publications Grant by the Royal Society from the Rockefeller Foundation Gift	300 0 0
200		<hr/>
		£5350 13 2

BALANCE SHEET AS ON 31 DECEMBER 1944

LIABILITIES		£	s	d	£	s	d
<i>Sundry Creditors</i>				172	18	10	
<i>Life Compositions</i>							
As on 31 December 1943		1609	0	0			
Add Payments during year		115	10	0			
		1724	10	0			
Less Transfer to Income and Expenditure Account		50	0	0	1674	10	0
<i>Subscriptions received in advance</i>							
Members		33	5	4			
Publications		315	19	9			
				349	5	1	
SPECIAL FUNDS							
<i>W. F. Stanley Trust Fund</i>		259	0	0			
<i>Duddell Memorial Trust Fund</i>		346	2	9			
"Progress Reports" Reserve Account		83	1	0			
<i>Herbert Spencer Legacy</i>		239	7	6			
<i>Charles Chree Medal and Prize Fund</i>		1949	3	0			
<i>Charles Vernon Boys Prize Fund</i>		907	1	7			
				3783	15	10	
GENERAL FUND							
As on 31 December 1943		5661	6	6			
Add Balance of Income and Expenditure Account		1144	1	2	6805	7	8

£12785 17 5

ASSETS		£	s	d	£	s	d
<i>Investments at Market Value on 31 December 1939</i>							
W. F. Stanley Trust Fund		259	0	0			
Duddell Memorial Fund		374	0	0			
Charles Chree Medal and Prize Fund		1866	10	0			
Charles Vernon Boys Prize Fund		900	0	0			
General Fund		6505	10	0			
				9905	0	0	
(Market Value on 31 December 1944 £11266)							
<i>Dividends due from Investments</i>							54 18 1
<i>Inland Revenue—Income Tax recoverable for 1944</i>							45 0 5
<i>Subscriptions due</i>							15 15 0
<i>Sundry Debtors</i>							277 15 7
<i>Stock of Paper and Binding Material</i>							199 6 6
<i>Cash at Bank</i>		392	16	2			
" in Post Office Savings Bank Account		1893	5	1			
" in hand		2	0	7			
				2288	1	10	

£12785 17 5

We have audited the above Balance Sheet and have obtained all the information and explanations we have required. We have verified the bank balances and the Investments. In our opinion such Balance Sheet is properly drawn up so as to exhibit a true and correct view of the state of the Society's affairs according to the best of our information and the explanations given to us and as shown by the books of the Society.

LIFE COMPOSITION FUND ON 31 DECEMBER 1944

20 Fellows paid £10	£	s	d
1 Fellow paid £10 10s.	200	0	0
1 Fellow paid £15	10	10	0
15 Fellows paid £21	15	0	0
36 Fellows paid £31 10s.	315	0	0
	1134	0	0
	<u>£1674 10 0</u>		

SPECIAL FUNDS

W. F STANLEY TRUST FUND

Balance Sheet	£	s	d	£300 Southern Railway Preferred Ordinary Stock	£	s	d
	259	0	0		199	0	0
				£442 Southern Railway Deferred Ordinary Stock	60	0	0
	<u>£259</u>	<u>0</u>	<u>0</u>		<u>£259</u>	<u>0</u>	<u>0</u>

DUDELL MEMORIAL TRUST FUND

CAPITAL															
balance Sheet		£		s.		d.		£400 3½% War Loan Inscribed "B" Account		£		s.		d.	
		<u>374</u>		<u>0</u>		<u>0</u>				<u>374</u>		<u>0</u>		<u>0</u>	
REVENUE															
31 December 1943		£		s.		d.		Interest on War Loan		£		s.		d.	
Certificates		36		5		3				14		0		0	
		5		12		0		Balance carried to Balance Sheet		27		17		3	
		<u>£41</u>		<u>17</u>		<u>3</u>				<u>£41</u>		<u>17</u>		<u>3</u>	

"PROGRESS REPORTS" RESERVE ACCOUNT

ried to Balance Sheet . . .	£	s	d	Balance on 31 December 1943 . . .	£	s	d
	83	1	0		83	1	0

HERBERT SPENCER LEGACY

ried to Balance Sheet . . .	£	s	d	Balance on 31 December 1943 . . .	£	s	d
	239	7	6		239	7	6

CHARLES CHREE MEDAL AND PRIZE FUND

CAPITAL									
ried to Balance Sheet	£	s	d.	Balance on 31 December 1943	£	s	d.		
	<u>1865</u>	<u>16</u>	<u>4</u>		<u>1865</u>	<u>16</u>	<u>4</u>		
REVENUE									
ried to Balance Sheet	£	s	d.	Balance on 31 December 1943	£	s	d.		
	83	6	8	Interest on Investments	14	9	0		
					68	17	2		
	<u>£83</u>	<u>6</u>	<u>8</u>		<u>£83</u>	<u>6</u>	<u>8</u>		

CHARLES VERNON BOYS PRIZE FUND

CAPITAL									
carried to Balance Sheet	£	s	d.		£1132 16s 10d 2½% Consols.		£	s	d.
	<u>900</u>	<u>0</u>	<u>0</u>				<u>900</u>	<u>0</u>	<u>0</u>
REVENUE									
carried to Balance Sheet	£	s	d.		Interest on Investment		£	s	d.
	<u>7</u>	<u>1</u>	<u>7</u>				<u>7</u>	<u>1</u>	<u>7</u>

THE HISTORY AND FUTURE OF THE PHYSICAL SOCIETY

BY PROF. E. N. DA C. ANDRADE, F.R.S.

Presidential Address delivered 23 May 1945

§ 1 INTRODUCTION

AT a time, like the present, of change and flux, at what is clearly the end of one great historic period and the beginning of a new phase of world history, it is appropriate that we should each examine our position and form appropriate plans for the future. What is true for individuals is no less true for corporate bodies, which have the double duty of satisfying the needs of the present and preparing for the demands of the future. It may, therefore, be appropriate for the retiring President to examine the present state of the Physical Society and tell the Fellows something of its plans for the future. Further, to judge the present position wisely, and to estimate our rate of progress, which, as well as our present position, it is necessary to know if we are to judge of our future—since positions and velocities are both needed if we are to calculate the future state of a system—it is advisable to make a brief survey of the past. I shall therefore run briefly through certain phases of the history of the Society, and refer to the introduction of the various innovations which now represent established features of the Society, so that we may see how our wise predecessors have built up the traditions which we now proudly possess. After some account of our present position, I propose to outline our ambitions for the future, with a word as to how we hope to realize them.

Before I proceed, however, I should like to refer to the present very healthy state of our Society and to give thanks to those to whom it is mainly due. Looking back to my own early days, I think that some of our younger brethren may be apt to take institutions with which they are connected—University departments, scientific societies, and so on—very much for granted. They may suppose that these exist, carry out their duties and serve by virtue of some inevitable and mysterious corporate existence, some appointed and predestined reason and source. The tides, the underground railway, the ten o'clock lecture, these have their being and their rules, but of how they are sustained and governed very few enquire in their youth. However, to conduct a society like ours, to ensure that your meetings are arranged and announced, to secure distinguished speakers and appropriate places of meeting, to see that the *Proceedings* are properly printed and delivered, and to maintain the efficient running of our organization in a hundred and one details is a matter that requires very skilled and devoted service, as well as considerable tact, knowledge and experience. In war-time, particularly, with shortage of staff, difficulties of communication, fire-watching and other duties not normally imposed, the running of the Society has been a very difficult matter, and we might well have found ourselves at this stage with our affairs in confusion, our publications in abeyance, our members apathetic and diminished, and our influence sadly lessened. That we are, on the other hand, in a flourishing state, with our membership steadily rising and our enthusiasm unchecked, is due in the main to our Honorary Secretaries. They have given far more than we had any right to expect, they have worked unsparingly and unceasingly for our good. Mr Awbery, the papers secretary, has, for instance, been his own editor of the *Proceedings*; Dr Jevons, our business secretary, has been his own assistant secretary. They, in fact, with a diminished office staff and a thousand difficulties, working hours which few of us would contemplate, have kept our flag flying in a way for which we can never be sufficiently grateful. I ask you to allow me, on your behalf, to offer our especial thanks to our loyal secretaries. Members of Council, too, have not spared themselves.

§ 2. BIRTH OF THE SOCIETY

Our Society was founded in 1874. Let us look back at the state of physics at about that date. In the preceding year, 1873, Maxwell had first published his *Electricity and Magnetism*, Crookes had invented his radiometer, and van der Waals had published

the papers containing his celebrated equation. Airy was completing his term of office as President of the Royal Society, and Lord Rayleigh, at the time the Hon John William Strutt, was elected a Fellow of the Royal Society his *Sound* was to appear four years later. J J Thomson was a student at Owens College, Manchester, under Osborne Reynolds and Balfour Stewart, of whom the last-named became our President in 1886.

In 1874, the year of our foundation, Tyndall was in control at the Royal Institution, but the greater part of his original scientific work had been completed. In that year, however, he delivered, as President of the British Association, his famous address at Belfast, which created a sensation and was, incidentally, the subject of a poetical summary by Clerk Maxwell, which summary was in its turn translated into Greek verse by R. Shilleto. Maxwell was at Cambridge, whither he had gone in 1873, although the main part of the researches by which his name lives was likewise completed. Faraday had died seven years before. Among the great contemporary names in physics were Kelvin, Stokes, Joule, Lockyer, Wheatstone and Crookes. van t'Hoff and Rayleigh were beginning to become famous: Hertz and Arrhenius were boys not yet heard of. In 1875 Oliver Lodge was appointed, at the age of twenty-four, reader in natural philosophy at Bedford College for Women. "As regards the general mechanical conveniences of life, the first typewriter was put on the market in this year. The gas engine had been realized and the petroleum internal combustion engine was in its infancy. The Atlantic cable had been in operation a few years. The telephone had been invented, but the first commercial telephone switchboard was not to come for a few years. There was, of course, no electric lighting in the home and no gas-mantle. The first gramophone did not appear until three years later. It may therefore seem hard to some to believe that people could spend happy lives in those days, but we are credibly informed that they did.

As regards the teaching of physics, the Cavendish Laboratory was completed in 1874 and instruction in practical physics started there under Clerk Maxwell. There was only one student in this year, W. M. Hicks, who later became Professor of Physics and the first Vice-Chancellor at Sheffield. Teaching of practical physics was going on at Manchester, under Balfour Stewart, with about ten students doing practical work in the laboratory. Carey Foster had started systematic practical instruction in physics at University College, London, in 1866, to be followed by Grylls Adams at King's College, London, in 1868, and Clifton at Oxford in 1872. Carey Foster had two rooms available for practical work, of which one was his private room and, as far as I can make out, there were about half-a-dozen students carrying out experiments. Carey Foster was the only one there, of course, carrying out research, and he first described his well-known bridge in 1872. As regards Oxford, there was no provision for research: in fact the advancement of knowledge was no part of the professor's duties, and Clifton took care not to exceed his duties. Rucker, the one demonstrator there, was succeeded in 1874 by W. N. Stocker, who became a Fellow of our Society in 1877 and is now our senior surviving Fellow. He recently made a most generous contribution to our Holweck fund.

It is perhaps worth noting that the first papers of Carey Foster, our second President, were all on organic chemistry. In particular, three papers which he published in co-operation with Matthiesen on narcotine are said, by those better able to judge than I am, to be a long step forward in the knowledge of the constitution of the alkaloids. J. H. Gladstone, our first President, subsequently became President of the Chemical Society in 1877, the only case, I believe, of one man having served as President of the two sister Societies. So close were the relations between chemistry and physics in those days.

Returning to the teaching of practical physics in 1874, we will note that it was proceeding under Kelvin in Glasgow, where in 1870 a new laboratory had replaced the disused wine cellars in which instruction was previously given, and then turn to South Kensington. Here, at the College of Science, which had been one of the results of the Great Exhibition of 1851, Guthrie was Professor of Physics, giving regular courses of instruction. Perhaps I may quote from a letter written by Silvanus Thompson in 1875: "I go at one p.m. to Professor Guthrie's lectures on Physics—at first on Hydrostatics and Pneumatics, but just now beginning *Sound*. He lectures every day except Saturday. He is a ponderous Scotchman, and puts in 'of course' about thirty times each lecture". (This is the kind of thing that students write about their professors!) Later, when he came to know Guthrie well, Thompson paid affectionate tribute to his services to science and to the Society. In the same letter Silvanus Thompson, who became a Fellow in

1875 and President in 1901, mentions that most of the students were teachers in training, and it must be remembered that almost the only future for a professional physicist in those days was some branch of the teaching profession. In particular, if you wished to carry out physical research you had either to hold a teaching post or to possess private means. As for the teaching of physics, and of science in general, in the schools, it was, according to a Royal Commission "regarded with jealousy by the staff, with contempt by the boys and with indifference by the parents". The general picture of British physics at the time is, then, one that shows a few intellectual giants, like Kelvin, Stokes and Clerk Maxwell, and a number of great but lesser known men keenly engaged in research, while laboratory instruction in physics was beginning at a few centres scattered over the country, at each of which half-a-dozen students or so laboured under some keen enthusiast who made up for indifferent accommodation and equipment by personal skill and devotion.

I have tried, very briefly, to indicate the state of physics as it was in England when our Society was born. The birth was due to a circular letter which Guthrie sent out, saying, "I wish to try to form a Society for Physical Research for showing new physical facts and new means of showing old ones, for making known new home and foreign physical discoveries, and for a better knowledge one of another of those given to physical work". The consequence of this letter was the inauguration of the Society in the Physical Laboratory at South Kensington, with Dr. J. H. Gladstone as the first President. Gladstone was at the time Fullerian Professor of Chemistry at the Royal Institution and is remembered for the work which he carried out with T. Pelham Dale on the refractive indices of liquids, resulting in their law connecting variation of refractive index with variation of density. The original Fellows who formed the Society numbered 99.

Professor Guthrie, who did not become President until ten years later, took the office of "demonstrator", which emphasizes that interest in experimental demonstrations which the Society has always shown, and still shows. No doubt the office corresponded to that of "curator of experiments", to which, in 1662, Robert Hooke was appointed in the newly formed Royal Society. It continued in the Society until 1897, when C. V. Boys, who had hitherto been "Demonstrator and Librarian", became simply "Librarian".

§ 3. SOME NOTABLE CONTRIBUTORS TO THE 'PROCEEDINGS'

Publication of the *Proceedings of the Physical Society of London* began at once, the first paper being one by J. A. Fleming, whom we lost this year, "On the new Contact Theory of the Galvanic Cell". In the same volume is a paper by McLeod, describing his celebrated vacuum gauge, a paper by Crookes on his radiometer—not the first description of it, but the first considerable discussion—and another by Guthrie on "Salt Solutions and Attached Water", giving an account of some of his classical work on cryohydrates. The names of Oliver Lodge and Carey Foster also occur. In the first ten years J. A. Fleming, Oliver Lodge, Norman Lockyer, Silvanus Thompson, R. T. Glazebrook, Ayrton and Perry, Arthur Schuster, J. H. Poynting and C. V. Boys are among the frequent contributors, and among contributions from America we have a considerable one by Hall, on the Hall effect—not the very first announcement of the effect, but following close upon it—and a note by H. A. Rowland on this discovery. The average yearly production for these first ten years is about 170 pages. Since those times the standard has been well maintained. In recent times such names as Finch, Ezer Griffiths, Laby, Lennard Jones, Lindemann, Mott, Raman, Rankine, O. W. Richardson, Rutherford and Chadwick and Temple adorn our pages, and I will in particular remind you that many of W. H. Bragg's papers first appeared in our *Proceedings*, and that Sir Edward Appleton's great series of papers on the ionosphere were likewise printed by us, beginning in 1928. One of his very earliest papers appeared in our *Proceedings* in 1921. He is a pretty shrewd judge, and chose the Physical Society for his papers, I believe, when he found out that not only is our standard high but that we can be relied upon to print important papers with great expedition and in a worthy manner. Of recent years our annual volume has consisted of some 700 or 800 pages, and has on occasion exceeded 1000 pages. We adopted our present large format in 1921, and I think it is generally agreed that our standard of production is commendable, so that as regards quality, quantity and presentation we have no reason to be ashamed of our *Proceedings*. We have, however, plans for extending our publication, to which I will refer later.

I have glanced at our *Proceedings*. another way of estimating our position in the world of physics is to look at the list of past Presidents. It includes Lord Kelvin, G. F. Fitzgerald, Sir Oliver Lodge, J. H. Poynting, H. L. Callendar, Sir Arthur Schuster, Sir J. J. Thomson, Sir Charles Boys, Sir William Bragg, and Sir Arthur Eddington, to mention only the dead. This is, indeed, a gratifying list, justifying our claim to have been supported—actively supported, for all these took a keen interest in the Society and did not regard the office as a sinecure—by the greatest physicists of the time. If I am to mention only one of the living past Presidents it shall be Sir Frank Smith, whom I choose because he presided brilliantly at our Jubilee Celebrations in 1924, and rendered outstanding services to the Society, both during and after his years of office.

§ 4 PUBLICATIONS

As regards our publications, in the first place we have our *Proceedings*, in which, as I have already said, most of the great names in British physics have appeared. From the time of Rowland we have from time to time been gladdened by American contributors. R. W. Wood in particular has sent us some of his most interesting papers.

In 1934 appeared the first volume of our *Reports on Progress in Physics*. It was J. J. Thomson, our President from 1914 to 1916, who first suggested the issue of these *Reports*, and he generously put up some money for a guarantee fund. Fortunately, after the first year this fund has not been called on. The *Reports* have found a warm welcome on both sides of the Atlantic and have evidently met a need. Volumes I to V are out of print. Four volumes have already appeared since the outbreak of war, and another, Volume X, is in active preparation. We have been fortunate in obtaining outstanding articles from American physicists, especially since the war. The fact that the energetic editor of the *Reports*, Dr. W. B. Mann, has been for some time resident in America has helped to secure American co-operation, which we greatly value. This may be a suitable place at which to express our thanks to the American Institute of Physics. This organization is not a counterpart of our Institute of Physics, but a publishing body which comprises the corporate membership of five societies, namely, the American Physical Society, Optical Society of America, Acoustical Society of America, Society of Rheology and the American Association of Physics Teachers. The Institute has been of great help to us during the war years by assisting in every way to distribute our *Progress Reports* among American physicists with the minimum formality, and has always shown us the greatest goodwill.

Besides these serial undertakings, we have from time to time issued special publications. Four special reports won instant recognition: those of Sir James Jeans on *Radiation and the Quantum Theory* (1st edition 1914, 2nd edition 1924); Sir Arthur Eddington on *Relativity Theory of Gravitation* (1920); A. Fowler on *Series in Line Spectra* (1922), and W. Jevons on *Band Spectra of Diatomic Molecules* (1932). Each of these appeared at a time when it filled an urgent demand. We published the *Collected Papers of Wheatstone* (1879) and of Joule (1884). The *Papers and Discussions* of the International Conference of Physics held in London in 1934, in the organization of which we played a great part, were published by the Society, and the proceedings of various Discussions organized by us, such as, e.g., the *Discussion on Vision* and the *Discussion on Audition*, have been published.

A word as to our future policy in regard to publications may be welcome. It was suggested at Council in 1943 that there was a need of a medium for the publication of original papers on physical problems whose main interest was industrial and technical, corresponding to the German *Zeitschrift für Technische Physik*, which was initiated in 1920, at the time of the foundation of the *Zeitschrift für Physik*. There seemed general agreement as to the reality of the need, and it has been decided, after due deliberation of the proposal by a Joint Committee set up by our Society and by the Institute of Physics, to which I shall refer later, that special provision should be made for papers on the more technical aspects of physics. It has not yet been decided whether to divide the *Proceedings* into two parts, called, say, "A" and "B", or whether to bring out a separate journal, to be called *Physica Technica*, or some such title. Here, again, Council will welcome expressions of opinion from the Fellows. In any case, everything points to the fact that it will soon be necessary for us to issue two volumes a year, instead of the customary one volume.

There is, however, a further publication in view. Many of the discussions of Groups are of great interest, while not describing in the first place original work. It is widely considered that there may be a place for a journal which shall deal mainly with discussions, papers which survey authoritatively certain restricted branches of physics, critical reviews of current problems and so on. Proposals have been made, but as nothing has yet been decided I can merely assure you that the matter is receiving consideration, and further, that nothing will be done by us without consultation, through the Joint Committee, with the Institute of Physics.

I cannot leave the subject of publications without referring to the goodwill that the Royal Society has extended of recent years to our *Proceedings*. I think that I may venture to say that any wish which we may cherish to take a greater part in the publication of British physical papers will not meet with a rebuff from that quarter.

§5 THE ANNUAL EXHIBITION

One of the most successful and useful undertakings of the Society is our Annual Exhibition of Scientific Instruments and Apparatus. This started from very humble beginnings. In 1905 there was held an "informal meeting", at which an exhibition of apparatus was given by the following firms: R & J Beck, Cambridge Scientific Instrument Co., Crompton & Co., Elliot Bros., Everett, Edgcombe & Co., Fricker & Miller, Peter Heale, Adam Hilger, Isenthal & Co., Marconi Wireless Telegraph Co., Nalder Bros. & Thompson, Newton & Co., R. W. Paul, Pitkin & Co., Rummey & Rummey, Synchronome Co., Carl Zeiss. Most of these firms are familiar to you and were among our most recent list of exhibitors. The attendance was about 240, which was greater than expected. In 1906 was held a like informal exhibition, with some fresh names, familiar to us now, among the exhibitors—I may mention Casella, Evershed and Vignoles, Gambrell, Ross, J. Swift & Sons, and Alexander Wright. In 1907 a more formal exhibition was held, which attained to the dignity of a notice in *Nature* (77, 159, 1907). Since that time the function has yearly increased in importance: at our last Exhibition before the war, the twenty-ninth, 80 firms exhibited, and the total attendance over the three days was nearly nine thousand. Our Exhibition catalogues of recent years have been most valuable guides to the state of the scientific instrument industry, not only in this country, but elsewhere, and the prudent have kept their copies. The catalogues as issued for the first few exhibitions were mere leaflets. I only know of one complete set of catalogues from the start, that in the possession of the Society.

It may be of interest to refer to one or two of the stages by which the Exhibition has progressed. In 1909 discourses of general interest, by leading physicists, were introduced as an attraction, the first two being given by C. V. Boys and Silvanus Thompson. In 1926 three new sections were added, devoted respectively to recent research of instrumental interest, effective lecture experiments, and apparatus; and experiments of historical interest. In 1930 a competition for craftsmanship and draughtsmanship was initiated, for learners and apprentices belonging to the exhibiting firms, which brought forth some exhibits of the greatest merit and awakened general interest.

As regards the future, you will be glad to hear that some months ago your Council decided that there was a good prospect that the Society would be able to hold an Exhibition in January 1946, and set up a Committee, under the chairmanship of Professor G. I. Finch, to organize matters. I am happy to say that the response from the industry is very encouraging, and we hope to have a display of instruments and apparatus that will be truly representative of the art and that will interest not only our own people but also to our friends from other parts of the world.

§6 PHYSICAL SOCIETY GROUPS

A recent innovation are our discussion groups, initiated during the war. Of these we have at present two, the Colour Group and the Optical Group. The Colour Group arose spontaneously out of discussion at one of our science meetings in 1940. At this meeting Dr. W. D. Wright suggested that the Society should undertake the organization of the many workers who were interested in the discussion of the various physical aspects of colours, so that they might have definite times and places of meetings and pre-arranged and pre-announced programmes of proceedings. Dr. Allan Ferguson, the reigning

President, who was in the Chair, favoured the suggestion, and as a result the inaugural meeting of the Colour Group was held in January 1941, under the Chairmanship of Dr. Wright, who very successfully conducted the Group through its difficult initial stages. In due course he was succeeded by Mr J. Guild, who recently handed over to Dr. R. K. Schofield a very flourishing organization, the present membership being about 160.

The Optical Group was founded, at the initiation of Professor L. C. Martin and Dr. W. Jevons, shortly after the foundation of the Colour Group. It was constituted to meet in the first place the specific needs of the members of the old Optical Society, which amalgamated with the Physical Society in 1932. Professor Martin was formerly an officer of the Optical Society. Under the wise guidance of its first chairman, Professor A. O. Rankine, who was our President from 1932 to 1934, the Group has held very successful war-time meetings, and now numbers about 350 members. The present Chairman is our old friend Instr.-Captain T. Y. Baker, R.N., who joined the Optical Society thirty-five years ago.

Shortly after our first Group was organized, the Institute of Physics started a policy of forming groups, and the inaugural meeting of its first group, which deals with *Industrial Radiology*, was held in May 1941. Two other groups were subsequently formed by the Institute, on *Electronics* and *X-Ray Analysis* respectively. I am glad to say that, with the object of avoiding competition and confusion in the setting-up of groups, in publications and in all matters in which both bodies are concerned, a permanent Joint Committee has been constituted "to facilitate co-operation between the Physical Society and the Institute of Physics". Some little time ago Sir Alfred Egerton wrote to me suggesting that the Society should form a Low-Temperature Group, for the discussion of the production and measurement of low temperatures and the properties of matter at such temperatures. In this connection low temperatures are understood to be those below -80°C or so and not only helium temperatures, which are, of course, included. It has now been decided, at the recommendation of the Joint Committee, to initiate such a Group, and we are calling an informal meeting of some of the leading workers in this field to discuss the matter. Any suggestions from Fellows will, of course, be welcomed by Council.

§ 7. ACCOMMODATION

Looking to the future, undoubtedly one of our greatest needs is suitable accommodation. Since 1939 we have, by the great kindness of The Royal Commission for the Exhibition of 1851, been housed at Lowther Gardens. Our sincere and cordial thanks are due to our generous hosts, the Commissioners, and in particular to their secretary, Dr. Evelyn Shaw, whose kindness and consideration will always be remembered with pleasure and gratitude by the officers of the Society. Without the hospitality of the Commissioners we should have been in sore straits. It would, however, be an affectation to pretend that the two uppermost floors of 1 Lowther Gardens constitute premises suited to our present needs. The rooms, originally bedrooms in a private house, can be adapted to ordinary office uses, but there is no room suited for use as a library, and, of course, no kind of a lecture room at all.

Our location in South Kensington is advantageous so long as we have to depend for theatres for our meetings on the kindness of the Imperial College, or on occasion of the Science Museum, whose director, Colonel Mackintosh, has always shown us great goodwill. It is, however, not very easy of access. We require a more central site and we require convenient rooms where we can not only carry out our office business and discharge our various duties to our Fellows, but also receive worthily our guests from the Empire and other countries overseas. Our needs may be said to be some eight adequate offices of one sort or another, suitable space for the storing of archives, an adequate library, a Council Room, a lecture theatre accommodating about a hundred auditors and a lecture theatre capable of seating some three hundred auditors. The Council Room and lecture theatres would, of course, be needed for intermittent, and not daily, use, and the economic scheme would be to share them with other scientific societies.

The ideal scheme of accommodation would, then, seem to be a central building adapted to house the chief scientific societies, a building in which each society would have its own block of offices, and in which council rooms, lecture theatres and public rooms would be at the disposal of the various participating societies. Common arrangements might advantageously be made for the printing of *Proceedings* and other periodical publications.

of the societies and for their dispatch. I am glad to say that our august mother, the Royal Society—for all the chief scientific societies in England were founded by her Fellows, and are in some sort descended from the Royal Society—is keenly interested in such a scheme. As the result of her activities a delegation, headed by the President of the Royal Society, was received on 13 October 1944, by the Lord President of the Council, the Chancellor of the Exchequer and the Minister of Works and Buildings. The object of the delegation was to represent the urgent need of a comprehensive scheme for the future housing of the Royal Society and other scientific societies. I was fortunate enough, in my capacity as your President, to be a member of this delegation and to be allowed to represent our needs to the Ministers present. Their reply to the delegation, although necessarily non-committal, was by no means discouraging. The Royal Society was asked to submit a detailed scheme of the needs of all the societies involved and has been collecting the information, of which we have supplied our part. It would be a graceful acknowledgment of the part which organized science has played in the successful prosecution of the war if the Government were to provide a worthy central home in which various scientific societies could live together under the fostering influence of the Royal Society.

Such a building would solve one of our most urgent needs, which is convenient and accessible library accommodation. As regards books, we are tolerably well off as far as the periodicals of physics are concerned, but we are less satisfactorily equipped with textbooks and specialized works. We have some historical books of interest, the gifts of Fellows of the Society. The library facilities are much needed by our Fellows, but the inaccessibility and inadequacy of our library space prevent full use being made of what books we have, and also stand in the way of development. I feel that if we could properly display our possessions and comfortably accommodate our readers, we should find that Fellows would support our library with gifts and bequests, and that, further, Council would allocate larger sums for library purposes. We have to welcome a very recent bequest from Mr G. L. Addenbrooke for the purchase of books of a special character, and we hope to receive further presents.

The matter of library organization under the central building scheme has already received attention. A meeting of librarians of various scientific societies, including our own, was convened by the Royal Society and deliberated on various possibilities. At this meeting we were represented by our Librarian, Professor L. C. Martin, to whose work for the Society I should like to pay tribute. I have not the time, nor have I obtained permission, to describe to you the results of these deliberations, but I thought that you would like to know that the whole matter of a central library covering various branches of science had been discussed from the point of view of peace, economy and convenience.

§ 8 SPECIAL LECTURES

We have under our control a series of special lectures of which we are proud. The oldest of such functions is the Thomas Young Oration, founded in 1907, which we took over when the Optical Society amalgamated with us in 1932. Of recent years this has been given biennially. We have often made this oration an opportunity of hearing one of our distinguished foreign colleagues, and this year, for the first time, we are to have an orator from Sweden, Professor Ragnar Granit, who will address us on colour perception.

The annual Guthrie Lecture was founded in 1914 to commemorate our founder, Professor Frederick Guthrie. The list of Guthrie lecturers includes some of the most famous names in physics. We have drawn our Guthrie lecturers from America, France, Denmark, Germany, Holland and Sweden, as well as, of course, Great Britain. Fellows will remember that last year we were delighted to draw, for the fourth time, our Guthrie lecturer from America, Professor Joel Hildebrand. This year we are to hear Professor A. Dupeyron, a Spaniard, who has been working in England for some years.

The biennial Charles Chree Medal and Prize was founded by Miss Jessie S. Chree to commemorate her brother, our President from 1908 to 1910, and carries a very handsome emolument. Professor S. Chapman, a Londoner, was the first recipient. Professor B. F. J. Schonland, from South Africa, the second: this year, as third Chree Medallist, we are to hear Dr J. A. Fleming, of the Department of Terrestrial Magnetism of the Carnegie Institution, Washington.

Of the Duddell Medal you have heard this afternoon. We shall shortly be proceeding to award a new prize, the Charles Vernon Boys Prize, which derives from a bequest of

money by Sir Charles Boys. It is to be given annually for distinguished research in experimental physics, carried out within five years preceding the date of the award.

The first Rutherford Memorial Lecture was given in 1942 by his colleague and our old friend, Professor H. R. Robinson. This year we are to hear the second Rutherford Lecture from Professor J. D. Cockcroft.

We have on hand a new project which I have very much at heart, the foundation of a Holweck Prize, the object of which is to foster friendship and co-operation between British and French physicists. It is named after Holweck, perhaps the most skilful experimenter that France has produced of recent years. He was the kind of man to whom we award the Duddell Medal, and he was murdered in Paris by the Gestapo. It is our intention that the prize shall be given in alternate years to a British and to a French physicist, and that each prize-winner shall visit the capital of the other country to receive it. We have appealed for subscriptions, and the initial response has been very encouraging. We are just inviting certain of the leading firms connected with physical research to subscribe to the Holweck fund, and I hope that all present will do their best to forward this project. The response of our French friends to our gesture has been all that we desired.

§ 9. THE SOCIETY TO-DAY

This leads me to say a word about our relations with other countries and with other societies. I have already referred to our very cordial friendship with the United States, from which country we so often draw our special lecturers. I may say that we have been promised some excellent articles from America for our *Progress Reports*, which have a warm welcome over there.

Last year we had the pleasure of welcoming M. Boutré, representing the Société Française de Physique, at a Science Meeting, and of greeting through him our French colleagues who have held up their heads and maintained their scientific and patriotic ideals through times of such hard trial, in spite of all that the brutal invader could do. We also heard an outstanding lecture from M. Guinier. Our relations with our French colleagues are warm and deep.

Within the Empire our position is firm. We have many Fellows in both Canada and Australia, nevertheless we should be very glad to welcome more. Professor Schonland, who recently went back to South Africa as the Director of Scientific Research for the Union, was for some time a member of our Council, and will, we are confident, uphold our interests there. Last year we were able to strengthen our ties with our Indian colleagues by taking part, both as a Society and as individuals, in welcoming the Indian scientists visiting this country as guests of the Government. Among them were two celebrated physicists, Professor Meghnad Saha and Professor S. K. Mitra, who addressed us on subjects to which they have made outstanding contributions, Professor Saha at a special meeting held in London, and Professor Mitra both in Cambridge and in London, choosing different subjects for his two lectures. We should like to see more Indian names recorded among our Fellows, and trust that one result of this visit will be an accession to our Indian representation.

In 1944 we elected Professor Joffé, of Moscow, to be one of our Honorary Fellows, the number of which is restricted to twelve. We hope that with improvement of communications we shall be able to hear and see more of our Russian colleagues. We are also keenly anxious to renew our former very cordial relations with the Scandinavian countries and with Holland and Belgium, and trust soon to see some of our old friends from those parts with us again.

While speaking of our international relations, I may say that the Presidents of the Société Française de Physique, of the American Physical Society, and of the Optical Society of America are *ex-officio* Fellows.

To our relations with the Royal Society I have already referred more than once. That great scientific Fellowship has clearly shown that it recognizes the substantial part that we play in the organization of British science and that it will further all our legitimate aspirations. We treasure the goodwill of the Royal Society, and respond with affection and gratitude. With the Faraday Society and the Royal Meteorological Society we are on cordial terms. We have shown how highly we think of the judgment of the latter by choosing one who was their President for three years as our own next President. I have

had occasion to speak of the Committee which has been set up to promote cordial co-operation with the Institute of Physics, and it is the wish of all of us to live on terms of close friendship with that guardian of the professional interests of British physicists.

A very strong feature of the Society is the body of Student Members. At present, by our Articles of Association, a student member must be between eighteen and twenty-six, and cannot remain a student for more than four years. He or she pays no entrance fee and a subscription of only half a guinea, which entitles to admission to all meetings and to receive the *Proceedings*, the Agenda Paper and Notices, and the Catalogue of the Exhibition. The object is, of course, to make it easy for the young physicist to take a full part in our activities and to attract him to become a Fellow when his four-year period is finished.

The Student Membership has increased very much of recent years. For the twenty years from 1903 to 1923 the average number of Student Members was about 7. In 1927 it reached 27, and then began a more or less steady increase. In 1944 the number was 279, the highest figure ever reached. About seventy per cent of our Student Members become Fellows.

We are, then, a flourishing Society, rejoicing in strong international friendships and cordial goodwill at home. That we are flourishing we may claim both in view of our scientific activities, which I have set before you, and of the growth of our membership, illustrated by the graph now shown. You will see that since 1917 our numbers have risen steeply and steadily, except for a bump beginning in 1932; this was due to the accession of members from the Optical Society, which merged with us in that year. We can claim, I think, to give our Fellows substantial benefits. The meetings which we organize and the very real fellowship which we foster there and elsewhere mean much to British physicists. We have a library, our ambitions for which I have touched upon when speaking of our aspirations in the matter of accommodation. In the way of publications we supply, free of charge, the *Proceedings*, the *Physics Abstracts* and the catalogue of the Instrument Exhibition, while Fellows can obtain the *Progress Reports* and all other publications at much reduced rates. Our publications and general activities, such as special lectures, have grown immensely since, say, 1910, yet our subscription has remained at two guineas since that date and earlier. In fact, the Physical Society has maintained pre-war—pre-1914 war—charges for a greatly increased service. I do not want to cast a gloom on this ardent gathering, but I feel that I ought to issue a warning that this state of affairs cannot reasonably be expected to go on for ever, and could not have persisted so long if we had not been running with an office staff much too small for our needs. We have managed to carry on only by the personal exertions of the Business Secretary, to which I have already paid tribute, but firstly he cannot be expected to continue in this way, and, secondly, he has, even so, only just been able to keep abreast of the essential work, with the consequence that there is a large accumulation of matters not of first urgency, but which now require attention. We must provide, and pay for, proper staff.

I have been talking of the state of our Society and of its standing in the world of science. I think that you will be interested to know that Council has decided that an application for the incorporation of the Society under Royal Charter shall be made when such a course becomes possible. During the war such incorporations have been in abeyance. If our application is granted we shall, of course, become the Royal Physical Society.

§ 10 THE FUTURE OF THE SOCIETY

Finally, I want to impress upon all Fellows that while, in ordinary parlance, they belong to the Society, actually the Society belongs to them. It depends for its vigour on the support that they give to it: its *Proceedings* are a reflection of their activities in pushing forward the bounds of science: the success of its meetings is measured by the interest which they display. Whether, as we confidently trust, the Society goes forward with ever-increasing usefulness, influence and lustre, or whether, as we have little fear, it declines to a dull mediocrity, is a matter that, in the end, depends on the Fellows. I can assure you, from detailed experience, that you have a Council made up of men with expert knowledge of physics, of physicists and of affairs, men who have the interests of the Society very near to their hearts and are willing to give up time and energy—nay, anxious to devote their best efforts—to promote these interests. They are not a set of

academic recluses or amiable figure-heads, desirous to see only their own friends on Council, but men who are eager to welcome representatives of the younger physicists when they show interest in the Society and the will and ability to help to advance its interests. I have, for instance, during my term of office as President, had the pleasure of welcoming to Council one who not so many years ago was a research student of mine. Council is always glad to receive from Fellows suggestions for the good of the Society and, in particular, recommendations as to new members of Council have always been and, I am sure, will always be, considered at election time.

These are times of great opportunity. We have been promised a greatly increased measure of financial support for the Physics Departments and their laboratories in our Universities. physics is being used in industry to an extent never before approached in England, and it seems to be realized in high circles that our future prosperity depends to a marked degree on the encouragement of physical research in its technical applications. We have an important part to play in fostering physical research in all its branches by affording worthy means of publication and by promoting discussion and friendly feelings among all physicists, whether engaged in teaching, pure research or industrial research. It is for us at our meetings to bring our younger Fellows and our Student Members into personal contact with the veterans of physics. We have important work to do in promoting international good feeling in the world of physics. I ask you to support your Society in an active manner, with something more than automatic co-operation—with vehemence, with enthusiasm. In your new President you have one who brings all the qualities necessary to lead our Society with brilliant success—outstanding eminence in one branch of our subject, recognized last year by the award of a Royal Medal—a wide knowledge of physics, both theoretical and experimental, to which his war work has added experience of men and things in various parts of the world: a genuine affection for the Society, which has received practical expression in the work which he has done for it as an active member of Council: a kindly, yet shrewd personality—and three years' experience of high office as President of the Royal Meteorological Society. He will lead you with skill and courage, with adroitness and with enthusiasm, through the many difficulties of the immediate future to a position of acknowledged power and of supreme use in the commonwealth of science. May our Society flourish and increase!



[Photo Elliott and Fry]

PROFESSOR E. N. DA C. ANDRADE, PH.D., D.Sc., F.R.S.,
President of the Physical Society, 1943-45

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 57, PART 1

1 January 1945

No. 319

THE INTERNAL RESISTANCE OF THE SELENIUM RECTIFIER PHOTOCCELL, WITH SPECIAL REFERENCE TO THE SPUTTERED METAL FILM

BY J. S. PRESTON AND G. W. GORDON SMITH,
National Physical Laboratory, Teddington

MS. received 13 September 1944

ABSTRACT. This paper describes the effect of the internal resistance (other than the barrier-layer) of the selenium rectifier photocell upon its observed characteristics. Methods of measuring the resistance of the selenium layer are cited. This resistance is of the order of 10 ohms in most modern cells. A method of measuring the resistance of the sputtered film is also described. This method is based on measurements, on unlacquered cells, of the distribution of potential over the film resulting from the flow of photocurrent across it when the cell is exposed to a steady uniform illumination. The results obtained with four cells tested gave values between 100 and 600 ohm cm. for the resistivity of the film. They suggest that for a cell of normal type the value is in the region of 100–300 ohm cm., while a value exceeding 500 or 600 ohm.cm. is likely to be associated with non-uniformity of the film over the area of the cell, and less satisfactory performance at high values of illumination.

The experiments were made on cells of a production type supplied by a regular maker

§1 INTRODUCTION

THE selenium rectifier photocell consists of a metal back-plate, usually of iron, coated on one side with a thin layer of light-sensitive crystalline selenium, on the front surface of which a very thin translucent metal film has been deposited, usually by cathodic sputtering. This film may be composed of a noble metal, but its exact nature in the commercial product is not ascertainable. While the cell is exposed to light it will maintain a current in an external circuit connecting the film to the back-plate, the only source of energy being the incident radiation itself. The photocurrent, regarded as a negative electron current, flows through the circuit from film to back-plate, so that the former is the negative electrode of the cell, and the latter is the positive. To facilitate electrical connection to the film, a narrow strip of fusible metal, called the contact strip, or ring, is sprayed on to the film near the periphery of the cell. The surface of the finished cell is usually lacquered to protect it from moisture and mechanical damage. The action of this type of photocell has been studied theoretically, but the studies have generally failed to give quantitative results in good accord with the measured characteristics. In part, this has been due to lack of knowledge of simple features of the cell, such as the resistance of its various elements, which are not easy to determine and may be subject to variation from cell to

cell. The present paper, dealing with the "ohmic" part of the cell resistance as distinct from the barrier layer, is therefore based principally upon experimental considerations. The object of the paper is to examine, simply, the influence of the internal resistance on the behaviour of the cell, and to form some estimate of the magnitude of this resistance. Particular attention is paid to the sputtered film, as it is believed that its resistance has not previously been measured *in situ* by a simple method such as is described below.

§2. THE ACTION OF THE CELL

While it is not proposed here to deal with the physical theory of the cell, it will be useful to have in mind a simple picture of its action. The primary action can be considered to be the liberation of electrons from the sensitive selenium layer by the incident light, the number liberated being proportional to the illumination. These electrons, in virtue of their being liberated with a certain initial energy, tend to accumulate on the sputtered film and charge it negatively, in spite of the opposing electric field so created. If there is no external connection between the film and the selenium layer, the film reaches a steady negative potential such that the flow of the photo-electrons on to it is just balanced by an internal back-leakage from the film to the selenium. The net current output of the cell is in this case clearly zero. If, however, an external circuit be connected to the cell, an extra return path for the electrons, from the film to the selenium, is provided. This reduces the potential difference between the film and the underlying selenium, and in consequence of the smaller opposing field, more of the primary photo-electrons flow into the film, within the cell, while fewer leak back. The inflow then exceeds the back-leakage, and the excess of the former over the latter provides the current in the external circuit. The lower we make the external resistance the lower we make the potential difference between film and selenium, i.e., the field opposing the collection of photo-electrons by the film, and so the greater becomes the current-yield of the cell for a given illumination.

Generally, for circuit resistances below a few hundred ohms, the resistance has little effect on the current output for a given (moderate) illumination. Almost all the primary photoelectric current then presumably flows into the external circuit, the opposing field within the cell being too small to cause appreciable back-leakage—or loss of sensitivity, as we may regard back-leakage.

We shall next proceed to a simple analysis of the action just outlined, showing how the internal as well as the external resistance plays a part in determining the potential difference between film and selenium, and so, also, the current output of the cell.

§3. THE RÔLE OF INTERNAL RESISTANCE IN THE BEHAVIOUR OF THE CELL

In the simple cell circuit mentioned above there are three parts, namely, (1) the interface or barrier layer, having rectifying properties, and assumed to be the seat of the conversion of radiant into electrical energy; (2) the remainder of the current path in the cell, having a resistance R_i , and (3) the external circuit

having a resistance R_E . Now if V be the voltage developed at the interface and i be the current flowing, we can at once write $V = i(R_I + R_E)$. We shall assume that V is the most important physical factor in this equation and that the higher the voltage V which the interface is called on to generate, the lower will be the sensitivity of the cell (defined as the ratio of the current to the incident illumination). It is seen that this is consistent with well-known experimental results. For instance, if we keep the circuit resistance constant and increase the current (by raising the illumination), or if we raise the circuit resistance and keep the current constant (by suitable adjustment of the illumination), or if we raise both circuit resistance and current output, the sensitivity is found to diminish. Our equation shows that in all three cases V has been increased. (It may be noticed here that the term iR_E on the R.H.S. of the equation is the voltage across the terminals of the cell).

For a proper insight into the behaviour of the cell, then, some knowledge of V is requisite. This at once involves R_I , although, of course, when R_E is large enough, it may often prove permissible to take the terminal voltage iR_E as a sufficiently good approximation to V , and neglect the term iR_I altogether. Even then, however, we must have some idea of the order of R_I at least, in order to know whether it is permissible to neglect it. At the other extreme, when the cell is short-circuited (or virtually so, as in the Campbell-Freeth circuit) and R_E is zero, R_I is the influential factor. We then have $V = iR_I$, and we see, for instance, that short-circuiting the cell does not make V (and so also the cell sensitivity) entirely independent of i , the current drawn from the cell. Short-circuiting the cell clearly reduces the range of variation of V , and with it the variation in sensitivity, for a given range of i , but it does not eliminate these variations altogether. (In other words, short-circuiting the cell results in better, but not exact, "linearity".)

The picture of the selenium photocell just given is admittedly incomplete, but it serves to show the importance of the rôle played by the internal resistance, and, together with knowledge of the approximate value, or even the order, of this resistance, it forms an adequate basis for very useful working rules in the use of the cells. For example, we may wish, for convenience in making certain observations, to assume that the cell sensitivity is constant, and to determine what conditions must be observed in order that the error resulting from this assumption shall not exceed a specified amount. We might then determine in the most convenient way (e.g. with R_E large and R_I certainly negligible by comparison) the appropriate value of V which must not be exceeded. Then, if we also know R_I roughly, we can insert the values of V and R_I in the simple equation given above, and so determine sufficiently closely for practical purposes the related ranges within which we must confine i and R_E .

§4 THE NATURE OF THE RESISTIVE ELEMENTS OF THE CELL

The purpose of this paper, in addition to drawing attention to the part played by the internal resistance in the behaviour of the cell, is to provide information on which a rough estimate of R_I for a typical cell can, if desired, be based, and to sketch the methods by which this information can be obtained for any given cell. We notice now, therefore, that the internal resistance comprises two

parts, namely, that of the selenium layer and that of the sputtered film. Methods of measuring the former are well known and are briefly referred to below. For the latter, a method is described which is believed to be sound and at the same time simple. Before describing these methods, however, an important point relating to the film must be referred to. We have so far assumed implicitly that V is uniform over the surface of the cell. Strictly, this assumption is not valid, for the resistance of the sputtered film leads to a non-uniformity in V . We ought, correctly, to apply the equation for V to each element of surface, rather than to the cell as a whole. R_I would then apply to a particular element under consideration, and the part of it due to the metal film only would clearly be smaller for elements near the contact strip than for remoter ones. This dependence of R_I for an element of surface, upon the location of the element, results in a variation of V over the surface of the cell. The variation may be large if either the resistivity of the sputtered film, or the current drawn from the cell, be unusually large. The consequences of such a case will be discussed later, for it is upon the variation of V over the cell surface that the present work is based.

§ 5 METHODS OF MEASURING THE RESISTANCE OF THE SELENIUM LAYER

As we have seen, the internal resistance of the cell, represented above by R_I , consists of two parts, namely, the resistance of the selenium layer and that of the sputtered film. The effective resistance of either will of course depend upon the current distribution in it. In the case of the selenium layer we may assume a uniform current distribution in almost all cases, since the sensitivity is generally uniform over the surface.

At the cost of destroying the cell it is possible to measure the resistance of the layer alone. For this purpose a comparatively heavy layer of fusible metal is sprayed on to the surface of the unlacquered cell, over the sputtered film, so as to form an upper electrode of negligible resistance. The cell in this condition will still have the properties of a rectifier. The effect of the interface having rectifying properties can, however, be eliminated by a well-known method. This consists in measuring the current passing through the cell for various values of voltage applied, in the "conducting" direction, from an external source. As the current is increased the differential of the current with regard to the voltage tends to a definite limit, while the resistance offered by the interface tends to zero. The limiting value of the differential is thus the resistance required. A close approximation to the limiting value can generally be obtained with applied voltages of 2 volts or less. Various investigators have used this method, and their results indicate that the resistance of the selenium layer in a cell of average size, say 45 mm. diameter, is generally of the order of 10 ohms.

A similar method might be applied to the cell in its original state, and expected to give a result representative of the total "ohmic" resistance of the cell, including that of the thin film, under normal conditions of use. It would not do so, however, because, as we shall see later, the resistance of the film is relatively high, so that if an external voltage were applied between the back-plate and the contact strip, there would be a concentration of current in the immediate neigh-

bourhood of the strip. Such a current distribution is quite unlike that of the photocurrent when the cell is used in the ordinary way. This method has sometimes been used in the past without recognition of its shortcomings

§ 6 MEASUREMENT OF THE RESISTIVITY OF THE FILM

It is possible, however, to measure the resistivity of the sputtered film alone. The method is based on the above-mentioned variation of V over the surface of the cell. The cell is exposed to a fairly high illumination and the variation in potential over the film is explored.* It is possible, with simple assumptions and for cells of simple shape, to relate the potential at any point on the film to the co-ordinates of that point, the film resistivity, and the photocurrent generated per unit area of the surface.

Consider for instance a rectangular photocell ABCD (figure 1), with the contact strips along the edges AB, CD. Assume that the properties of the sputtered film, and also the cell sensitivity, are uniform over the surface. Then, if the cell be uniformly illuminated, the current flow over the film will be symmetrical with regard to the centre line drawn parallel to AB, CD. The electron current will, in fact, flow everywhere away from the centre line, and in a

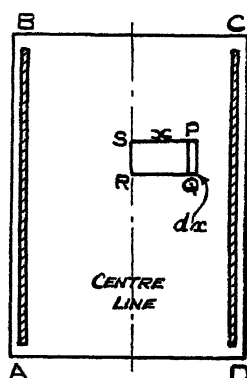


Figure 1

direction at right angles to it. Thus, all the current arising within a rectangular transverse area PQRS, where RS is along the centre line, will flow across PQ. Now let PQ be of unit length, and SP be x . Also let c be the photocurrent generated per unit area of the surface, and R the resistance, between two opposite sides, of a square of film of side unity. Consider then the small rectangular element of film whose longer dimension is PQ and shorter one dx . The current flowing across this element will be the current generated within PQRS, which is cx . The resistance of the element, in the direction of the current flow, will be Rdx . Hence the potential difference dp , across the element, will be $cRxdx$. If p be the potential difference between points on the line PQ and the centre line we have by integration

$$p = cRx^2/2. \quad \dots\dots (1)$$

* Unlacquered cells are necessary. Specimens were kindly supplied by Messrs Evans Electro-selenium, Ltd., to which firm due acknowledgment is made

(For a circular cell with an annular contact strip, the corresponding relation is easily seen to be $p = cRr^2/4$, where r is the distance from the centre of the cell)

We may test whether a relation of this form holds for a given cell, and if it does we can measure p , c and x , and so deduce R . The value of c is of course obtained as the quotient of the total photocurrent by the effective sensitive area. The following account describes the details of such a test.

A rectangular cell was mounted upon the movable system of a travelling microscope so as to be capable of movement in its own plane in a direction at right angles to the two contact strips (figure 2). A probe, made of fine springy

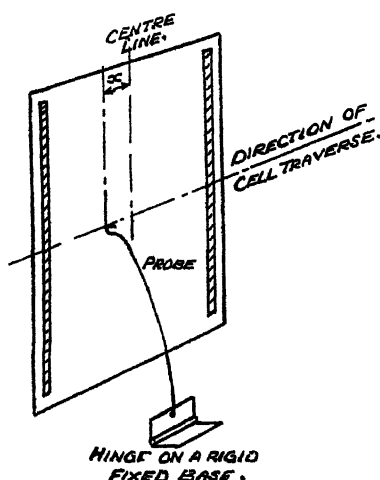


Figure 2

wire, with the end turned back on a small radius, so as to give a smooth point of contact, was mounted on a hinge fixed on the stationary frame of the microscope. The probe could thus be brought into contact with the cell, or removed at will while the position of the cell was being changed, without uncertainty as to the relative positions of the cell and probe. The micrometer screw of the microscope movement was used to vary, and to measure accurately, the position of the cell relative to the probe, and contact with the cell surface was thus possible at any desired point on a transverse line across the cell. The central transverse line was, in fact, chosen so as to avoid possible "end-effects". The cell was exposed to a steady uniform illumination, and by the means just described, readings of the potential of the cell surface were made on the transverse line, at points spaced 2 mm. apart. A potentiometer method was employed, so that no current flowed into or out of the cell, via the probe, at the moment of measurement. Instead of measuring the potential difference, p , between the probe and the centre of the cell, at which no permanent form of contact could easily be provided, measurements were made between the probe and a conductor connecting the two contact strips together, values of p being obtained as differences. The conductor joining the contact strips was also connected to the negative terminal of a 15-ohm micro-ammeter whose positive terminal was connected to the back-plate. This instrument measured the total photocurrent.

The internal resistance of the selenium rectifier photocell

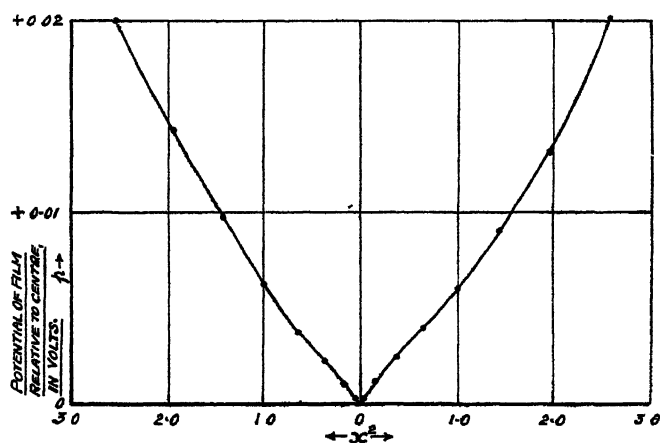


Figure 3 Square of distance from centre line, in cm^2

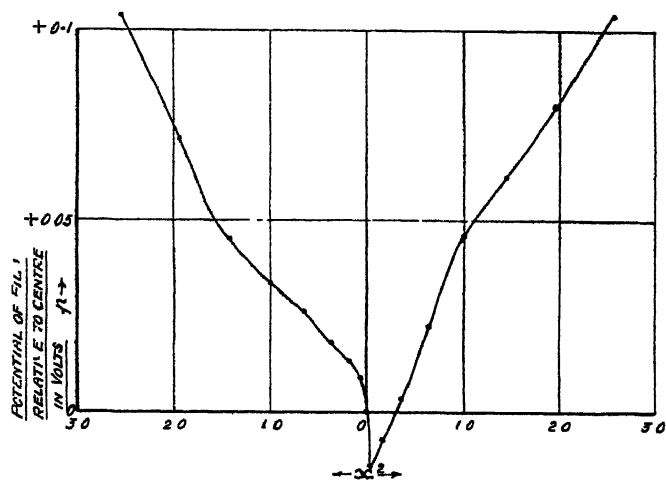


Figure 4 Square of distance from centre line, in cm^2

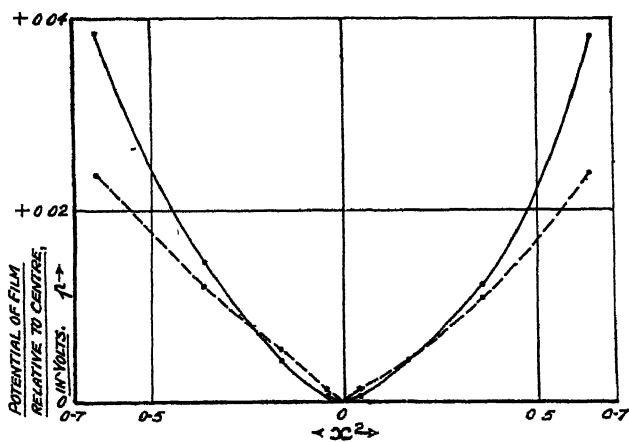


Figure 5. Square of distance from centre line, in cm^2

Care was taken that the illumination was uniform over a sufficient area, and that the total photocurrent was maintained constant for any one set of readings corresponding to a complete traverse of the cell.

§ 7. RESULTS

The curves shown in figures 3, 4 and 5 show some typical results. The corresponding data are given in table 1. Readings were taken in every case for several different values of the total photocurrent less than those stated, but the resulting curves were all of the same shape as those shown. The values of p observed at any particular point on the cell were, in fact, proportional to the total photocurrent, to within a few per cent.

Cells nos. 1 and 3 were of regular production types, while nos. 2 and 4 were sputtered with rather thinner films as a matter of interest. The higher sensitivities of the latter are consistent with a smaller light-absorption in the film. For the measurement of sensitivity the illumination was 10 f.c. and the resistance of the microammeter was 10 ohms. The light source had a colour temperature of 2848° K.

Table 1

Cell no	Cell size (mm)	Sensitive area (cm ²)	Sensitivity* ($\mu\text{A./f.c.}$)	Total current (mA.)	Current per unit area ($\mu\text{A./cm}^2$)	Potential diff centre to edge (volts)
1 (Fig. 3)	37 × 50	15.0	7.0	2.0	133	0.0201
2 (Fig. 4)	37 × 50	15.0	7.3	2.0	133	0.104
3. (Fig. 5 broken curve)	22 × 40	5.9	2.8	1.5	254	0.0238
4. (Fig. 5 full curve)	22 × 40	5.9	3.0	1.5	254	0.0385

(* refers to the unlacquered cell.)

In the case of cell no. 2 the observed relation between p and x^2 departs considerably from the linear form. The possible reasons for this are discussed later. For cells nos. 1 and 3 the relation approximates fairly closely to the linear form, while the results for no. 4 are intermediate. We shall for the present

Table 2

Cell no	x^2 (cm ²)	p (volts)	i (A./cm ²)	R (ohm cm.)
1	2.56	0.0201	$1.33 \cdot 10^{-1}$	118
2	2.56	0.104	$1.33 \cdot 10^{-1}$	611
3	0.64	0.0238	$2.54 \cdot 10^{-1}$	293
4	0.64	0.0385	$2.54 \cdot 10^{-1}$	474

disregard the departure from linearity and calculate the value of R in each case from equation (1) by inserting the value of p corresponding to the potential difference between the centre and the contact strips or edge of the cell and the value of x^2 for the half-width of the cell. This will presumably give a rough average value of R in cases where the film is non-uniform. Table 2 shows the values of resistivity so obtained.

§ 8 DISCUSSION OF RESULTS

The results indicate that the resistivity of the sputtered film, for an average normal cell, lies in the region 100–300 ohm cm. Cells with a thinner film than usual had a resistivity of 500 or 600 ohm.cm. These had a higher sensitivity (table 1), at least at an illumination of 10 foot-candles, no doubt because the thinner film absorbs less light. It is seen, however, that the film resistivity must be made two or three times higher than usual in order to secure so small an increase in sensitivity as 7 %. (It is, of course, well known that for very thin sputtered films the resistance decreases very rapidly with increase in thickness, while the transmission factor falls only slightly over the same range in thickness.)

Consider now the curves showing the observed values of p . These can be regarded as curves showing the variation of V over the surface of the cell (though the ordinates are not actual values of V , since V includes any potential drop in the selenium layer and also the terminal voltage). Figure 4 (cell no. 2), then, illustrates a case where the value of V at the centre of the cell must be at least 0.1 volt greater (arithmetically) than the value at the edge. So that even if the cell were short-circuited, and the peripheral regions were called on to furnish only a small voltage, the central region would be generating a voltage of at least 0.1 volt. This difference in V between the centre and the edge of the film is associated with the passage of the photocurrent through the film on its way to the contact strip. The difference is, of course, greater the greater the film resistivity, and increases with the photocurrent. It is easy to see, then, that when a cell having a poor film of unusually high resistance is called on to furnish a large photocurrent, V may tend to reach a value much exceeding 0.1 volt near the centre of the cell. Now for most commercial cells a rather marked fall in sensitivity sets in for values of V in excess of about 0.1 volt, while for cells of poor quality the critical value of V may be much lower than this. In the conditions just considered, therefore, the sensitivity of the central regions of the cell will be considerably lower than that of regions near the contact strip—that is to say, the sensitivity of the cell as a whole will be lower than it would be for much smaller photocurrents, for which V nowhere reaches such a high value.

Hence, too thin a film may be detrimental to the sensitivity of the cell as a whole at high illuminations, though it may result in a higher sensitivity than usual at low illuminations because of its comparatively high optical transmission. This is presumably one reason why the film on a normal cell is, as we have seen, made a little thicker than is necessary to give maximum sensitivity in the region of 10 foot-candles illumination. Sensitivity at low illuminations is not, for these reasons at least, a reliable measure of the general "goodness" of a cell.

It has frequently been suggested that, in order to avoid the conditions in which the film might be damaged or V become excessive over certain areas of the cell when a large photocurrent is being generated, extra central contact strips might be provided. Such a device is, however, of academic rather than practical interest, for the use of several small cells of standard design, in parallel, provides an equally satisfactory and rather more convenient solution. Probably, also, the variation in potential over the surface of the cell accounts for the peculiarity often observed with cells prepared by those not thoroughly versed in the art, that the sensitive portion appears to be confined mainly to regions near the contact strips. The film resistance in such cells may be abnormally high, and this would result in a serious reduction in sensitivity per unit area in the central regions, even for quite small values of the photocurrent. Moreover, as we have seen, in a poorly made cell the critical value of V may be much lower than 0.1 volt, and so the effect of a high film resistance would be further accentuated.

Let us now discuss the possible reasons for the observed departure from linearity of the relation between p and x^2 in the case of cells such as nos. 2 and 4. It may arise from non-uniformity in the sensitivity of the *selenium* over the surface of the cells. Now we might well suppose that a cell which shows good uniformity does so because the maximum possible sensitivity has been secured for every element of the surface. If so, its overall sensitivity must be higher than that of a non-uniform cell in which by definition the sensitivity cannot everywhere be a maximum. However, the cell having the anomalous characteristic of figure 4 is rather more sensitive than the one to which figure 3 applies, and which certainly exhibits good uniformity. It seems highly improbable therefore that the peculiar shape of the curve in figure 4 arises from non-uniformity in the state of the selenium. A similar argument applies, but with less cogency, to the cells of figure 5.

It is much more likely that the observed departures from the linear relation result from non-uniformity in the resistance of the sputtered films, with a distribution of photocurrent different from the simple one assumed in deriving equation (1). This is in accord with the observation already made that variations in the thickness of a film which have only a very small effect on the amount of light transmitted (and therefore on the sensitivity of the cell) may be associated with large variations in resistance.

The method described above for measuring the resistivity of the sputtered film on a selenium rectifier photocell would seem, then, to be applicable with confidence and to give the desired result to a good order of accuracy. When the sputtered film is non-uniform, this fact is immediately revealed by the method, but an approximate estimate of the average resistivity may still be obtainable. The method stresses the misleading nature of a simplification often made in describing the cell and its mode of action. This simplification is to regard the film, as a whole, as having a certain resistance through which the whole photocurrent flows. To do this is strictly incorrect, though it is often permissible in considering the general behaviour of the cell, as in the earlier part of the present paper. It might be urged that direct measurements of the resistivity of sputtered films deposited on glass, for instance, would have yielded all the information desired. This is not so, for the character of a sputtered film depends on the

material on which it is deposited, so that films sputtered on glass and upon selenium under identical conditions may not be alike in their properties.

§9 CONCLUSIONS

We may briefly summarize our investigations into the behaviour of the selenium rectifier cell as follows —

(1) A simple study of the part played by the internal resistance of the selenium rectifier cell serves to show the main features in its influence on the behaviour of the cell. A knowledge of the approximate value of the internal resistance is useful in this respect, and also as a guide to the experimental conditions appropriate to any practical application. The internal resistance comprises both the resistance of the selenium layer and that of the sputtered film. These two differ in many important ways, and must be considered and measured separately.

(2) The only convenient and valid method of measuring separately the resistance of the selenium layer involves destruction of the cell. Other investigators have obtained values of the order of 10 ohms by this method.

(3) The method now described for measuring separately the resistivity of the sputtered film is based on an exploration of the potential distribution over the film under normal conditions of use. The cell is in no way injured. The distribution is found to be of a type predicted on theoretical grounds. Values of 100 to 300 ohm.cm. are found for the resistivity of the film on cells of a regular production type. Other cells tested gave results of 500 to 600 ohm.cm. for the mean resistivity, with evidence that when the resistivity is as high as this, the film may not be very uniform, and the performance of the cells in certain respects unsatisfactory.

(4) It is correct to regard the cell as being made up of a large number of small elements, the effective film-resistance of each element depending on its position with regard to the contact strip. The simplified picture in which the film is regarded as offering some definite resistance to the photocurrent as a whole is frequently misleading. No value of resistance which might correspond to such a picture is therefore given. A value appropriate to any size of cell and set of conditions may be calculated, however, when this simplification is justifiable.

ACKNOWLEDGMENT

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

A NOTE ON THE SPECTRUM OF MgO

BY R. F. BARROW AND D. V. CRAWFORD,
Physical Chemistry Laboratory, Oxford

*MS received 13 October 1944, in replacement of a
preliminary note received 13 May 1944*

ABSTRACT Some observations on the spectrum of magnesium burning in air have been made with a grating spectrograph of dispersion $\sim 7.4 \text{ \AA/mm}$. Wave-length data are given for a weak sequence of bands with head at about 4820 \AA and of a complex system in the region $3600\text{--}4000 \text{ \AA}$. from measurements of plates taken in a first and second order respectively. The $\lambda 4820$ sequence may possibly form part of the well-known green system of MgO. The stronger bands of the near ultra-violet system have also been photographed in absorption.

§ 1. INTRODUCTION

THE most prominent features in the spectrum of magnesium burning in air are the well-known band system in the green, a weaker system at longer wave-lengths, a complex series of bands apparently degraded in both directions in the near ultra-violet ($3600\text{--}4000 \text{ \AA}$), and the $\lambda 5211$ system of MgH, all superimposed on a continuous background of radiation from the incandescent oxide. Although our primary concern has been with other aspects of this spectrum, we have found that the description given in the literature of the green system and of the bands at $3600\text{--}4000 \text{ \AA}$. is inadequate in certain respects. The purpose of this note is to provide some further data which may be of use in any future and more fundamental work.

§ 2. EXPERIMENTAL

Small strips of magnesium ribbon were burned in air and the light focused directly on the slit of a 2.4-m. grating instrument. Most of the observations were made in a first order (linear dispersion about 7.4 \AA/mm), but the bands at $3600\text{--}4000 \text{ \AA}$. were also photographed in a second order. Ilford Ordinary and Special Rapid Panchromatic plates were used in the appropriate wave-length regions. they were developed in Ilford Process developer to secure as much contrast as possible.

§ 3. THE GREEN SYSTEM

A reproduction of a grating spectrogram of the green system is given in the plate (a). The strongest features are seen to be two sequences of closely-spaced bands degraded to the violet with leading heads at about 5206 \AA . and 5007 \AA .: the former is somewhat overlaid by MgH structure. These bands have been measured many times, and the approximate wave-lengths of the earlier members of the two sequences are not in doubt. However, a third, weaker sequence, with head at about 4820 \AA ., is also evident. This sequence was apparently

first observed by Brooks (1909), who gave rough measurements to $\pm 1 \text{ \AA}$. The only other measurements known to the authors are from the low-dispersion work of Ghosh, Mahanti and Mukkerjee (1930), although the bands are also to be seen on the reproduction given by Eder (1904). Our figures are given in table 1

Table 1

$\lambda (\text{\AA})$	$\nu_{\text{vac}} (\text{cm}^{-1})$
4825.62*	20717.0
18 55	747.4
10 15	783.6
01.57	820.7
4791.24	865.6
80 45	912.7
70 75	955.2

* This band is considerably weaker than the others, it can scarcely be seen on the reproduction (see plate (a)), but seemed rather more definite on the negatives. A band at about this wave-length is listed both by Brooks and by Ghosh, Mahanti and Mukkerjee

The appearance and occurrence of these bands at once suggest that they represent a third sequence of the green system of MgO, and it is in fact possible to include them in a Deslandres array which numerically is quite feasible. With the $\lambda 5206$ band as the 0,0 band, the band at 5007 \AA . becomes the first member of the 1,0 sequence, and the bands given in table 1 fit satisfactorily into the positions of the 2,0 sequence. This scheme leads to values of the vibration frequencies as follows: $\omega'_i \sim 760$, $\omega''_i \sim 710 \text{ cm}^{-1}$.

There is a serious objection to this arrangement, however, which is based on the distribution of intensity among the bands. Although most photographic plates show rather sharp changes of sensitivity around 5000 \AA ., it is probable that the sequences do decrease in intensity in the order $\lambda 5007$, $\lambda 5206$, $\lambda 4820$, as indicated in the plate (a). On a rough visual intensity scale of 10 we would guess the figures 10, 6 and 1 respectively. The above scheme therefore represents a rather anomalous intensity distribution with the 1,0 band as the strongest. The only comparable distribution in the literature appears to be that given by Meggers (1933) for an infra-red system of CaO ($\nu_0 = 9491 \text{ cm}^{-1}$). This system consists of three sequences of bands degraded sharply to the violet, and the intensity figures listed are: (0,0) band, intensity 3; (1,0), 30, (2,0), 20. So far as we know, in all other systems arising from states with similar τ_e and ω_e values, leading to narrow Condon parabola, the strongest band is invariably the 0,0 band (see Jevons, 1932).

On intensity grounds, therefore, the present analysis must be regarded as rather improbable. However, the only obvious way of reconciling this analysis with a normal intensity distribution would appear to involve rejection of the $\lambda 5206$ band—a procedure for which there is no other justification. In these circumstances it is clear that further work is necessary to decide the nature of the $\lambda 4820$ sequence: a search for the isotope effect in the $\lambda 5007$ and $\lambda 5206$ bands on high dispersion, using a source free from MgH emission, would probably be decisive.

§4. THE ULTRA-VIOLET BANDS

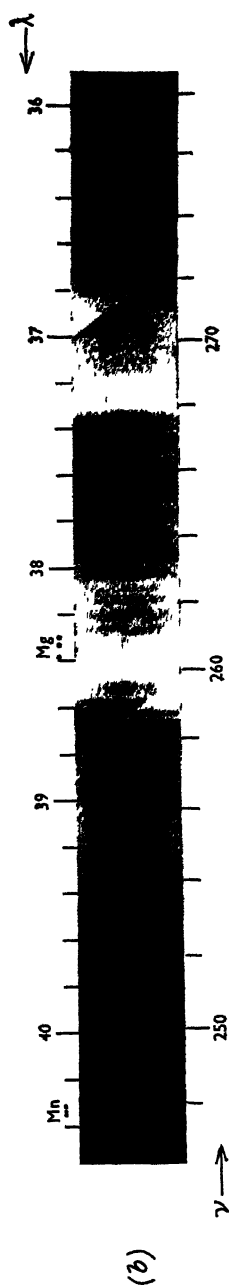
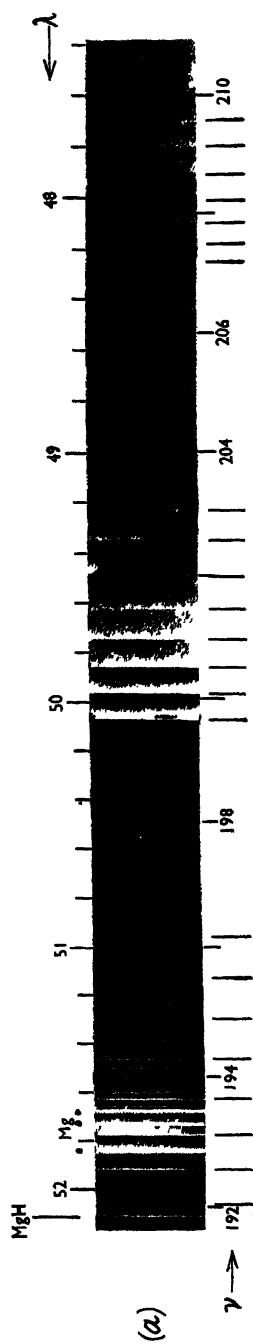
The complexity of the ultra-violet band system appearing in the spectrum of magnesium burning in air will be apparent from the reproduction given in the plate (b). Much of the presumably vibrational structure appears to be incompletely resolved even at a dispersion of 3.6 Å./mm., as on our second-order

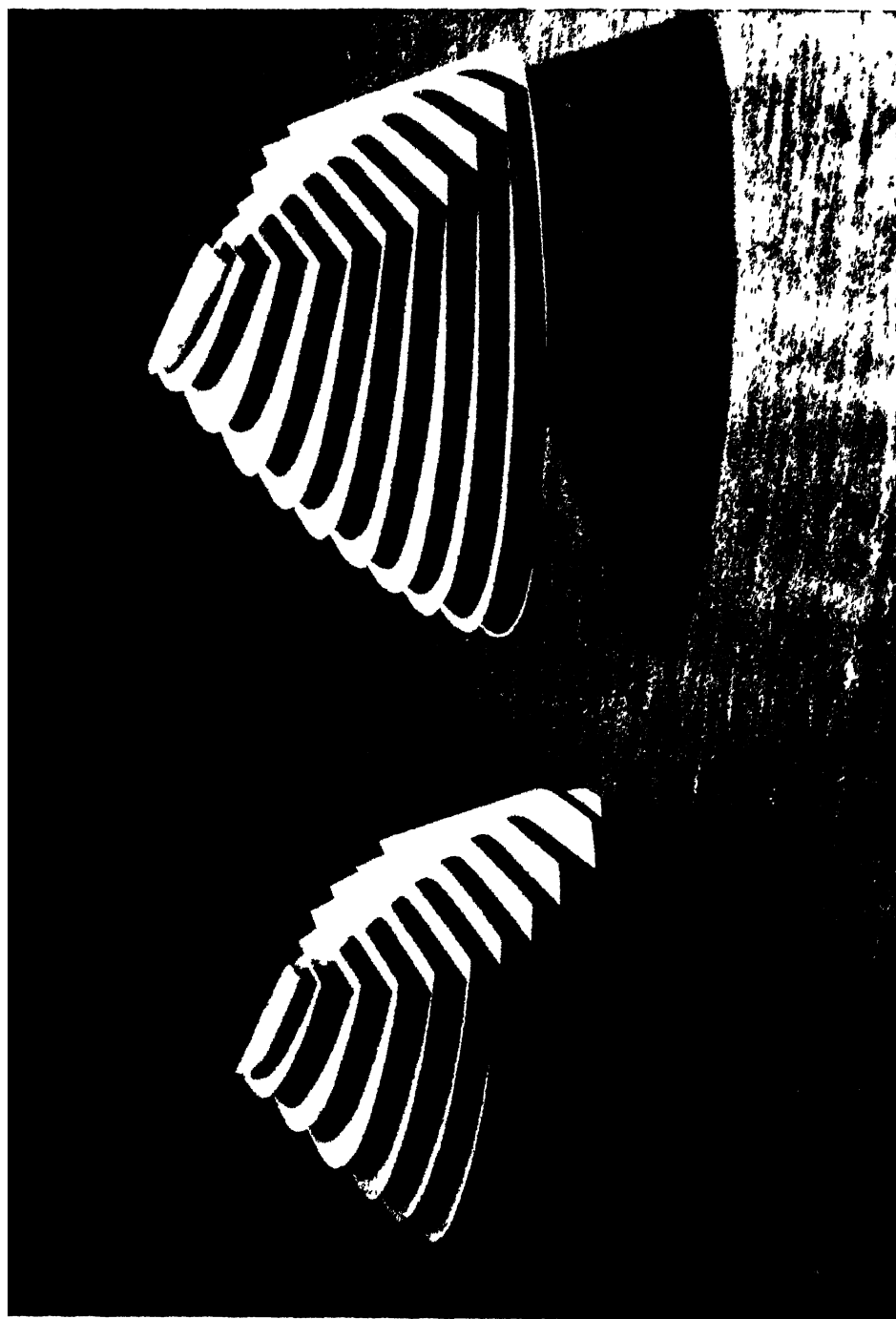
Table 2 Ultra-violet bands in the spectrum of magnesium burning in air

λ (Å)		Int	ν_{vac}	λ (Å)		Int	ν_{vac}	
3594.23	V	0	27814.5	3772.8	R	A	3	26498
99.4	V	0	774.5	73.05	V	A	3	496.2
3612.1	V	0	677	77.37	R		2	466.0
21.25	V	1	607	77.83	R		2	462.7
26.80	V	1	564.7	82.25	R	A	2	431.8
28.4	V	1	552.5	83.33	V		2	424.3
33.86	V	A	511.1	83.95	R		2	419.9
39.7	V	1	467	84.48	V		2	416.3
60.12	R	2	313.8	90.65	R	A	2	373.2
62.5	R	1	296	92.58	V	A	2	359.8
64.5	R	1	281	98.29	R		4	320.2
65.1	V	1	277	3802.6	V	A	2	290
71.9	R	1	226	04.23	R		3	279.1
72.5	V	1	222	05.28	R	A	5	271.8
74.73	R	2	205.2	10.28	R	A	5	237.4
84.0	R	A	137	15.60	R	A	5	200.8
84.92	V	A	129.9	16.5	R	A	5	195
89.07	V	A	099.4	23.8	R	A	5	145
90.61	R	2	088.1	24.21	V	A	5	141.8
98.64	V	2	029.3	35.74	V		5	063.2
3700.8	R	2	014	42.30	V		6	018.7
02.6	R	2	000	45.12	V		7	25999.6
03.78	M	A	26991.8	48.48	V	A	8	976.9
09.30	V	A	951.6	49.77	V	A	8	968.2
14.6	V	5	913	52.46	V	A	4	950.1
20.71	V	A	869.0	53.49	V	A	5	943.2
21.04	V	A	866.6	54.98	V	A	5	933.2
21.36	V	A	864.3	56.32	V	A	6	924.1
21.83	M	A	860.9	59.17	V		4	905.0
24.82	V	A	839.3	60.1	V	A	4	899
25.86	V	A	831.8	61.68	V		4	888.2
26.03	V	A	830.6	65.02	V		4	865.8
31.4	V	A	792	65.49	V		4	862.6
31.92	V	A	788.3	76.88	V		1	786.7
32.26	V	A	785.8	80.18	V		1	764.7
40.65	V	2	725.7	82.43	V		3	749.8
41.42	R	2	720.3	3906.5	R		1	591
50.61	R	A	654.8	13.0	R		2	549
55.6	R	A	619	21.25	V		1	494.9
59.42	V	A	592.3	27.6	R		2	454
66.00	R	A	545.9	36.17	M		2	398.5
66.4	V	A	543	46.0	V		2	335
69.7	V	A	520	52.2	V		1	295
69.98	V	A	517.8	63.1	V		1	226
71.79	R	2	505.1	71.4	V		1	173
72.39	R	3	500.9	94.5	V		0	027

The letters R, V indicate band degraded to red or violet.

M indicates a maximum, and A denotes a band observed also in absorption.





a McAdam colour solid

Figure 2 (*a* and *b*)

b Modified colour solid

plates. The bands seem to form very closely spaced sequences, initially degraded to shorter wave-lengths, but the presence of bands degraded in the opposite direction suggests that the sequence difference may change sign at high ν' , ν'' values. All attempts at a vibrational analysis have failed, and we conclude that work with higher dispersion is necessary.

The identity of the carrier must remain in doubt until such work has been done, but we would point out that the system is of considerable interest, particularly in view of the fact that we have observed it in absorption in the flames of certain pyrotechnic compositions containing magnesium.

The only reliable measurements of the main features of the system are those of Eder (1904), but his data apparently refer to maxima rather than to probable heads of bands. We therefore list wave-lengths and rough intensity figures of the more prominent heads in table 2.

REFERENCES

- BROOKS, E. E., 1909 *Astrophys. J.* **29**, 177
EDER, J. M., 1904 *Denkschriften der kaiserlichen Akademie der Wissenschaften Mathematisch-Naturwissenschaftliche Klasse, Wien*, **74**, 45
GHOSH, P. N., MAHANTI, P. C. and MUKHERJEE, B. C., 1930 *Phys. Rev.* **35**, 1491
JEVONS, W., 1932 *Report on Band Spectra of Diatomic Molecules* (Physical Society), p. 69
MAHANTI, P. C., 1932. *Phys. Rev.* **42**, 609.
MEGGERS, W. F., 1933 *Bur. Stand. J. Res., Wash.*, **10**, 669

THE BRIGHTNESS OF PRESENT-DAY DYES

By T. VICKERSTAFF,

Dyehouse Department, Imperial Chemical Industries, Ltd.,
Dyestuffs Division, Blackley, Manchester, 9

Lecture given to the Colour Group 22 January 1944

§1 INTRODUCTION

IN dye research laboratories, new dyes are continually being evolved, some of which ultimately become commercial products. Those which are so developed must clearly display advantages over the dyes already on the market, since the dye user would not otherwise change over from established products. The advantages shown by a new dye may be of many types, such as greater resistance to light or washing or easier application than existing dyes, but the aspect of greatest interest to the Colour Group is that of the colour of the dye. A new dye may be marketed solely because of its novel colour or outstanding brilliance, and it is obviously of great practical interest to the dye manufacturer to know just how the colour qualities of a dye may be improved and, in particular, the extent to which improvement in any given direction is possible, so that research may be directed towards the most promising field.

Improvement in the colour qualities of a dye may be obtained in two directions, namely, novelty of hue and increased "brightness". It is desirable that at the outset of the present paper the significance of the word *brightness* should be

defined. The dyer uses the word in its everyday sense, as when one refers to a "bright red coat", or a "bright green hat", and it is with this meaning that it will be employed here. The physicist, unfortunately, uses the same word in a quantitative manner, referring to the amount of light reflected from a surface. In order to avoid confusion it is proposed to use the term *lightness* to cover this latter usage. The *brightness* of the dyer covers a combination of *lightness* and *purity*. Thus if a comparison is being made of two samples of equal hue and equal purity, then the lighter will be termed the brighter, while if the patterns are of equal lightness but different purity, then the purer pattern will be termed the brighter. This matter has been discussed in some detail in an earlier paper (White, Vickerstaff and Waters, 1943).

Returning to the question of possible colour improvement, the main objective of the dye manufacturer is the production of greater brightness than is shown by existing dyes, the reason being that bright dyes can always be made duller if necessary whereas dull dyes cannot be brightened. There is not so much scope for the invention of dyes of novel hue, since there are already sufficient dyes to form a satisfactory colour circle, and intermediate hues can be obtained by mixture. In certain cases there may be advantages in having a homogeneous dye of a particular shade since it may be brighter than a mixture, and in any case a single dye is usually more easy to apply to material. There is some gap, for instance, in the yellow-green region of the hue circle.

Nevertheless, the predominant importance of brightness is fully realized in the colour-using trade, and dye manufacturers are constantly striving to produce brighter reds, brighter greens, brighter blues and brighter yellows than those already in existence. It is obvious that this process cannot be continued indefinitely. There must be a limit to the brightness which it is possible to attain in any particular hue. It may be that some existing dyes already approach that limit closely, so that attempts to improve them would be useless, while there may be other regions in which attempts to improve brightness would have much greater scope. It is clearly not possible to decide these matters by a mere inspection of the available colours, as there can be no innate knowledge of the attainable limit. However, the limits of purity and lightness which may be attained by surface colours have been defined by McAdam (1935), and it would seem a simple matter to compare real dyes with these limits and thus determine the hues offering most scope for improvement. It was with this object in mind that the present work was undertaken. The comparison was found to be a more difficult matter than appeared at first sight and although a fairly satisfactory basis of comparison was finally worked out, it is felt that the difficulties encountered are in many ways as interesting as the final conclusions, and it is therefore proposed to give a brief chronological account of the approach to this problem.

§ 2 BASIC IDEAS

As an introduction to the subject, the reasons for the existence of limits to the purity and lightness attainable may be briefly outlined. Any coloured surface can be defined in terms of the trichromatic or the monochromatic system, and such a definition is complete. The three variables involved are usually regarded as being completely independent, it being possible to vary any one whilst the

others remain constant. While this is true when applied to coloured light, it is not entirely true of surface colours such as painted panels or dyed fabrics. In such coloured surfaces, the effect of colour is produced by the absorption of part of the incident light, and all such surfaces must therefore have a lightness of less than 100%, since they are measured by comparison with a standard white surface which reflects all the incident light. Furthermore, it is clear that as more dye

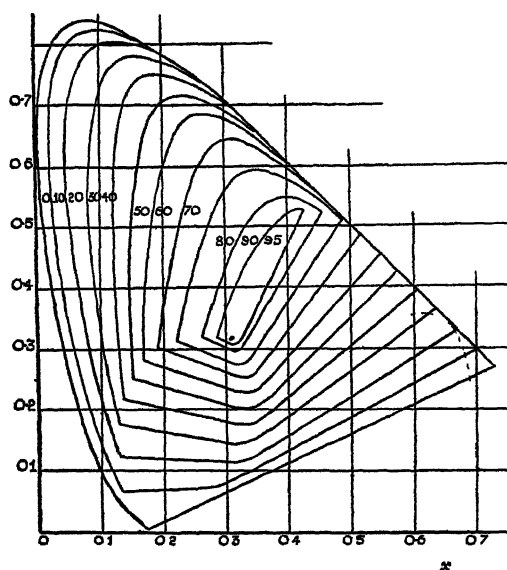


Figure 1 a.

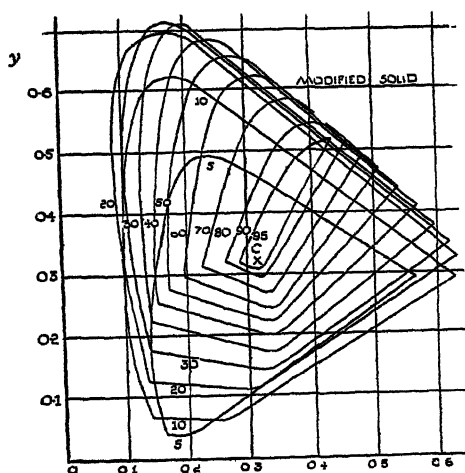


Figure 1 b.

is applied to a surface, which thus differs more and more from white, so does the colour of the surface become purer, but at the same time darker. High purity and high lightness are thus mutually exclusive in the case of surface colours, and to this extent these two properties are not independent variables.

This question has been investigated in detail by McAdam, who has calculated In order to do this, McAdam first proved that the coloured surface will have maximum purity and lightness if the spectral reflection curve of the surface satisfies the following two conditions:

- (1) The reflection factor at any wave-length must be either unity or zero.
- (2) There must not be more than two transitions in the value of the reflection factor over the visible spectrum.

The application of these conditions leads to the conclusion that coloured surfaces possessing maximum purity and lightness must have rectangular-shaped spectral-reflection curves having not more than two absorption bands in the visible region. Curves of this type may be converted by normal methods into the corresponding trichromatic or monochromatic specifications, and thus it is possible to calculate the maximum possible lightness which a coloured surface can exhibit at any given hue and purity. The most convenient representation

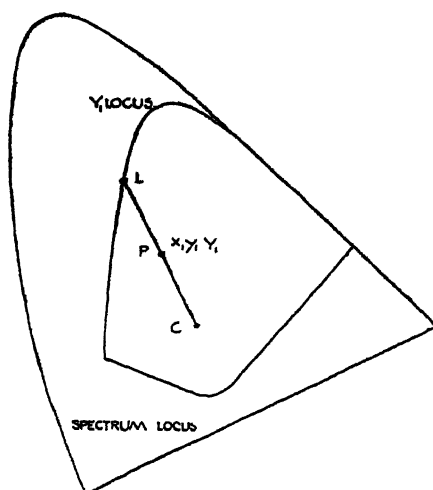


Figure 3.

of all the results is in the form of a colour solid within which all real coloured surfaces must lie. A contour map (after McAdam) of this colour solid is given in figure 1 *a* and a photograph in figure 2 *a*. The base of the solid is the x, y plane and the vertical ordinate is the lightness.

The limits of this colour solid provide the yardstick against which the achievements of real dyes must be measured. The difficulty which immediately arises is the exact manner in which the measurement is to be made. In a first attempt to overcome this difficulty, it was proposed to express the approach of a real dye to the ideal limits (its "efficiency") merely as the ratio of the excitation purity of the real sample to the excitation purity of the limit of the colour solid at the same hue and lightness. The method is illustrated in figure 3. A certain pattern has co-ordinates x_1, y_1 , and lightness Y_1 , and is, therefore, located at the point P. A cross-section of the colour solid at the same lightness level Y_1 has the contour shown, and the point P really lies in the plane of this cross-section. If C

the maximum possible purity of a coloured surface of any given hue and lightness. represents the position of illuminant C and if the line CP cuts the limit contour at L, then it was proposed to define the efficiency (E_s) of pattern P as

$$E_s = \frac{CP}{CL} \times 100 = \frac{e_P}{e_L} \times 100,$$

where e_P and e_L are the excitation purities of points P and L respectively.

The efficiency of a single dyed pattern can be expressed in this way without much difficulty, but, unfortunately, the efficiency of a dye varies with the depth of shade employed. As the amount of dye on a pattern increases, so does the purity up to a certain point. If more dye is placed on the pattern after this point has been reached then the purity usually decreases. The dyer expresses this by saying that the pattern becomes duller at high concentrations. The efficiency of dyed patterns when each pattern is compared against its own limiting point usually behaves in a similar manner, rising to a peak and falling again. For this reason it was decided to define the excitation efficiency of a dye as the maximum efficiency which could be obtained on dyed fabric at any depth of shade, and in order to determine this it was necessary to measure a whole range of dyeings of gradually increasing strengths for each dye. The dyeing having maximum efficiency was then selected as representing the efficiency of the dye itself.

§ 3. EXPLORATORY EXAMINATION

As an example of the method and the type of variation encountered, the complete results for one dye, Victoria Pure Blue BOS, are given in table 1. In the

Table 1. Excitation efficiency of dyeings of Victoria Pure Blue BOS

% dye on wool	x	y	Actual lightness	Excitation purity	Dominant wave-length	Theoretical lightness	Excitation efficiency
			%		A	%	%
0.06	189	194	16.7	58.0	4774	20.9	68
0.125	175	149	11.0	69.0	4733	13.8	78
0.25	165	113	6.6	78.5	4696	8.3	82
0.50	168	093	4.4	81.2	4655	5.5	85
1.00	172	088	3.3	81.0	4630	4.1	85
2.00	183	082	2.6	79.8	4570	3.3	85
3.00	190	092	2.1	76.0	4540	2.6	79
4.00	201	099	2.0	72.0	4480	2.5	75

first column of the table is the percentage of dye applied to the cloth, and in the next three columns the trichromatic co-ordinates and the lightness of the resulting dyeings in Illuminant C as measured on the Donaldson colorimeter. Then follows the dominant wave-length and excitation purity obtained by plotting the co-ordinates on the graphs given in Hardy's *Handbook of Colorimetry*, with Illuminant C as white light. The penultimate column gives the approximate lightness which the patterns would have if the substrate were perfectly reflecting. Undyed woollen cloth of the type used in these investigations is slightly yellowish and has a lightness of 80% (mean of several measurements). Since McAdam's

data are calculated on the assumption that the reflectance of an uncoloured surface is 100%, it is necessary in assessing the efficiency of a dye to make some correction for the absorption of the substrate, by dividing the actual lightness values by 0.8 an approximate value is obtained for the lightness of a dyeing on a perfect substrate. The error introduced by the fact that the undyed wool is not unselective in absorption is very small. Finally, the efficiency of each pattern is shown, from which it is clear that the maximum excitation efficiency attained by this particular dye is 85%.

The method was empirical, but it did provide a first basis for a comparison between dyes of different hues. Working on this basis, a number of dyes were examined. The dyes were of various types and included acid dyes on wool, vat dyes on cotton and pigments in oil painted on glass. Efficiencies varying from 40% to over 90% were obtained, and it was found that there was a general consistency in the results on the different types of dye, greens having low efficiency, and yellows high efficiency. The variation of efficiency with colour is shown in

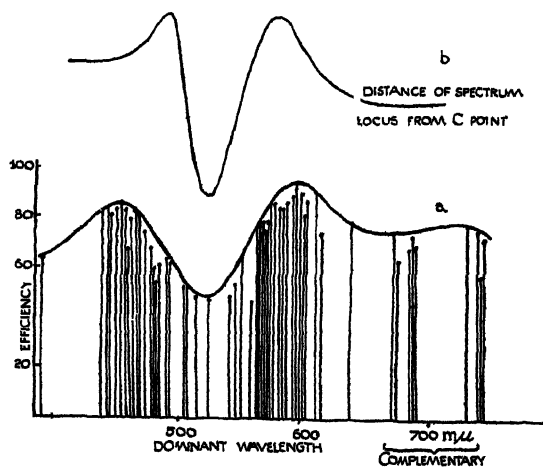


Figure 4 (a and b).

figure 4, in which the efficiency of 61 dyes or pigments is plotted against the dominant wave-length of the respective colours. All the colours tested have efficiencies falling within the area bounded by the smooth continuous curve, which has two particularly characteristic regions, namely that of very high efficiency near 600 mμ (yellows and oranges) and that of low efficiency near 520 mμ (greens). Examination of this diagram indicates that the points of maximum and minimum efficiency are located at wave-lengths corresponding to the points at which the spectrum locus is respectively nearest to and furthest from the point representing Illuminant C. This immediately suggests that efficiency is in some way determined by the geometry of the chromaticity diagram, and in fact the effect may be traced to the use of excitation purity, or distance in measuring efficiency. Excitation purity was chosen for this purpose because it is easier to determine than colorimetric purity, and it was thought that a comparison of the excitation purity of a real surface with the theoretical maximum excitation purity at the same lightness would be a sufficient measure of efficiency, and would indicate

the limits which can theoretically be obtained, just as well as the more complex colorimetric efficiency. Further consideration shows, however, that although the limits are indicated in this way, the scale of efficiency is not uniform over the spectrum, so that comparison of dyes of differing hue is not justified.

The concept of efficiency would appear to be placed upon a sounder basis if colorimetric purity is employed, as this takes due regard of the fact that the relative "brightness" of the spectrum varies with wave-length.

§ 4. REVISED FORMULATION

Accordingly, the previous results were recalculated to give the colorimetric efficiency E_c according to the definition

$$E_c = \frac{C_P}{C_L} \times 100,$$

where C_P and C_L are the colorimetric purities of P and L respectively. On this basis, there seems some justification for assuming that the results will be independent of hue.

The colorimetric efficiencies are plotted against dominant wave-length in figure 5 *a*, from which it will be seen that the efficiency now attains a maximum

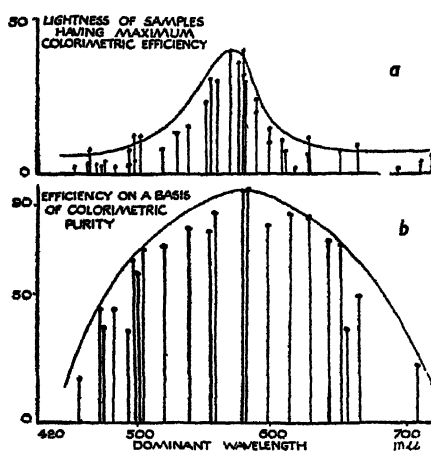


Figure 5 (*a* and *b*).

value in the green and falls away steadily on either side. After some consideration it was found that if the lightness of the patterns giving maximum purity is plotted against dominant wave length (figure 5 *b*) the curve shows a maximum in the same position as the efficiency curve. Thus the efficiency must be related in some way to the lightness of the patterns. This behaviour was finally attributed to the light scattered from the front surface of the patterns. When light falls on any surface, a portion is reflected without penetrating the material, and, therefore, consists of unchanged white light. The remainder of the light passes into the material, suffers partial absorption and emerges as coloured light. The fraction of light reflected from the front surface is practically independent of the colour of the dye, and on the average has a value of about 2% of the incident light for most textile materials, although it is modified by many factors, such as

the surface structure of the dyed material and the refractive index of the medium. If the efficiency of a pattern is measured at a relatively high lightness, then the addition of 2% white light will have little effect upon the colorimetric purity of the pattern, but if it is measured at a lightness of, say, only 5%, then the white surface light will comprise 40% of all the light coming from the pattern and will reduce the colorimetric purity very greatly, so that the efficiency will also decrease. The variation of colorimetric efficiency with wave-length shown in figure 5 is due, therefore, to the presence of this surface light, and an attempt was made to correct for this effect.

One of the conditions laid down by McAdam in his derivation of the colour solid which is the basis of the present discussion was that the reflection factor at any wave-length must be either zero or unity. With any real coloured surface this condition cannot be attained, for even if a dyed material did absorb completely all the light at some wave-lengths, its over-all reflection factor would still be 2% owing to the surface scattering. In order to correct for this, a new colour solid was constructed on the assumption that the reflection factor at all wave-lengths, for patterns having maximum purity and lightness, is either 1 or 0.02.

Table 2. Trichromatic co-ordinates of contours of maximum purity at constant lightness for surfaces reflecting a minimum of 2% of the incident light

5% lightness		10% lightness		20% lightness		30% lightness		40% lightness	
<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>
0.570	0.292	0.634	0.294	0.634	0.325	0.623	0.348	0.600	0.376
440	378	500	405	500	445	500	461	400	562
340	447	400	483	400	536	400	552	300	642
260	496	300	561	300	621	300	637	240	686
240	502	200	614	240	667	240	681	220	700
220	502	200	621	200	690	200	704	200	707
200	489	180	624	180	698	180	710	180	702
180	452	160	622	160	699	160	709	160	678
165	400	140	609	140	692	140	698	138	600
148	300	125	580	120	679	111	600	128	500
138	200	108	500	100	642	104	500	128	400
140	150	100	400	091	600	106	400	132	300
145	100	100	350	086	500	118	300	138	220
159	040	101	300	092	400	134	176	160	215
180	040	108	200	104	300	160	172	200	204
200	042	117	140	120	200	200	164	280	183
210	044	127	100	134	126	260	154	320	173
220	050	139	068	160	124	300	146	330	172
260	076	160	066	200	120	320	144	340	174
400	175	180	064	240	116	340	149	360	184
		200	060	280	112	360	160	400	214
		240	059	300	113	500	260	500	296
		260	064	320	120				
		280	074	360	143				
		300	087	500	236				
		450	180						

Table 2 (continued)

50% lightness		60% lightness		70% lightness		80% lightness		90% lightness		95% lightness	
<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>	<i>x</i>	<i>y</i>
0 571	0 408	0 538	0 442	0 510	0 470	0 483	0 496	0 455	0 523	0 430	0 520
400	572	400	573	400	574	400	574	420	596	400	508
320	639	320	638	360	606	380	584	400	547	360	470
300	654	300	649	340	617	360	588	380	540	320	400
260	680	280	660	320	625	340	582	340	500	288	320
240	688	260	660	300	625	314	560	292	400	308	308
220	687	240	652	280	616	279	500	264	318	314	307
200	672	211	600	265	600	246	400	300	301	320	309
172	600	187	500	225	500	224	312	320	296	328	320
156	500	175	400	207	400	252	300	330	304	369	400
150	400	172	300	196	300	320	271	400	422		
150	300	172	280	240	280	340	277				
150	253	200	268	320	248	400	368				
200	237	300	232	330	246						
300	208	320	226	340	248						
320	202	330	224	360	270						
330	201	340	225	440	377						
340	203	360	240								
360	212	460	353								
400	248										
500	341										

The actual value of the first surface reflection will differ according to whether wool, cotton or glass plates are examined, but in order to avoid the labour of computing a fresh solid for each surface, a value of 2% for the scattered light was assumed throughout. This is probably near enough to the real value in the case of textile fabrics, since the lightness of black dyeings rarely falls below this figure. In the case of glass plates coated with pigments the value is probably a little high.

The new solid was constructed by the method described by McAdam, using tables of running sums for the separate trichromatic functions constructed from the tables in Hardy's *Handbook of Colorimetry*. The resulting colour solid is shown in figures 1*b* and 2*b* for comparison with the unmodified solid, and the marked reduction in the possibilities of high excitation purity in the greens may be noted. The co-ordinates of several points on the loci at various lightness levels are given in table 2.

The efficiency of the patterns was now redetermined on the basis of this new colour solid, again comparing the colorimetric purity of each pattern with that of the theoretical limits at the same lightness. When this comparison was carried out at different dye concentrations, it was found that the efficiency of each dye increased steadily with increasing concentration and showed no maximum. This might indeed have been anticipated from the behaviour of dye solutions. If the concentration of dye in a solution is increased, the purity of the transmitted light increases continuously and ultimately reaches 100% if the dye is sufficiently

soluble. This is because all dyes must have a maximum transmission of light somewhere in the spectrum, and, at a sufficient concentration, the light transmitted will therefore tend to become monochromatic. In the case of dyed patterns, however, the behaviour is different, the purity increasing to a maximum value as the concentration increases, and then decreasing, this effect is due to the uncoloured surface-scattered light. Consequently, if a correction is applied for this factor, then the efficiency of dyed patterns must also approach 100% at very high concentrations.

Clearly, then, while the concept of colorimetric efficiency can be applied to any particular pattern or surface, it cannot be used to attach any value to a specific dye, since it varies with concentration. It is possible to overcome this difficulty by specifying that the efficiency must be measured at some particular lightness or purity, or even dye concentration, but such methods are purely arbitrary, and would give unequal weight to colours of different hue. It may also be mentioned that the alternative method of measuring efficiency, as the percentage of the limiting lightness which a pattern attains at the same hue and purity, is open to similar objections, namely, that efficiency varies with concentration, and that to select an arbitrary purity at which efficiency should be measured would be to discriminate against certain hues. Many other possibilities were considered, but all suffered from one or more of the defects of arbitrariness, non-uniformity of scale, or variation in value with concentration. Another method which was investigated, in a tentative manner, was to compare the shape of the absorption curves obtained by plotting optical density of a solution against wave-length with the corresponding rectangular curves postulated by McAdam. The advantage of this approach is that the shape of the curve is independent of concentration; but there are many drawbacks, such as the fact that efficiency measured in this way is a purely empirical function which has no bearing on the visual effect, and that, with the apparatus available, practical measurements were limited to solutions, whereas interest chiefly lies with dyed materials, and in many cases the colour of a dye on the fibre may be different from its colour in solution.

§ 5. FINAL METHOD

At this stage the problem was reviewed in an attempt to define more clearly the aim of the investigation. There seemed to be several objectives possible, but after careful consideration it was decided to concentrate attention on trying to find a means for measuring the possibilities of improvement in the colour of a dye in terms of sensation steps rather than in physical units. The reason for this is purely practical, for the dye-maker is interested only in the visual sensation produced by a certain improvement in colour, and not in its physical dimensions. The most satisfactory approach appeared to consist in a determination of the number of perceptible steps of change by which a given colour differs from the ideal McAdam limits. It might appear at first sight that this could be ascertained in a fairly satisfactory manner by plotting the points upon a uniform chromaticity diagram of the type suggested by Judd (1935) or by Breckenridge and Schraub (1939). Unfortunately, such diagrams are designed to apply only to lights of nearly equal intensity, and cannot be applied to questions involving surface colours, which necessarily vary very considerably in lightness. The

only satisfactory solution is to measure the distance between the real dye and the ideal limit inside a uniform chromaticity colour solid, in which, by definition, unit distance in any direction corresponds to an equal difference in colour sensation, and which is based upon the known behaviour of surface colours.

The only available information on a colour solid of this type is upon the Munsell colour system, the patterns of which have been specified in terms of trichromatic units in recent publications (Kelly, Gibson and Nickerson, 1943).

In the Munsell system, the colour attributes of a pattern are expressed in terms of the three quantities *hue*, *chroma* and *value*. These correspond qualitatively to the *hue*, *purity* and *lightness*, respectively, of the monochromatic system, but quantitatively the relationship is complex and cannot be mathematically defined, since the monochromatic system is an expression of colour in terms of physical quantities, whereas the Munsell system represents an attempt to arrange colours in such a manner that the visual spacing appears uniform and regular. In view of the recent extensive smoothing of the spacing of the Munsell colours, based on visual assessments (Newhall, Nickerson and Judd,

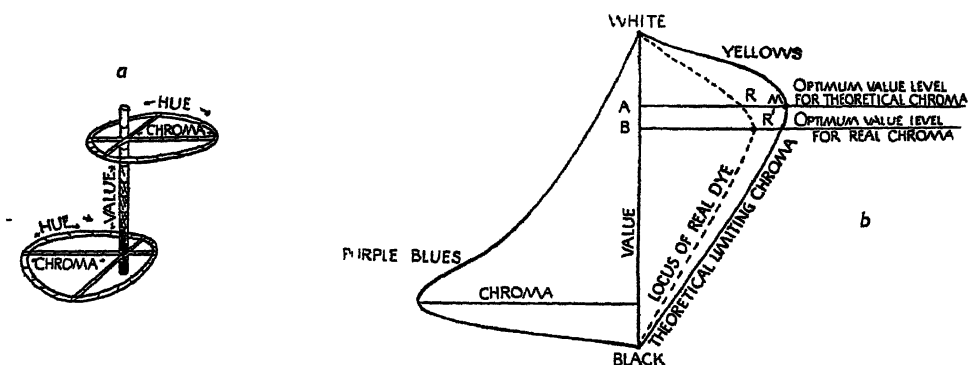


Figure 6 (a and b).

1943), there is a reasonable prospect that a unit difference of chroma at any value (lightness), and in any hue, will produce a similar visual or mental effect. This system thus appears to offer the greatest promise of providing a method whereby the visual possibilities of colour improvement can be assessed in a way which will be independent of the hue and lightness of the patterns examined, as far as uniformity of scale is concerned.

The possibilities of improvement, however, will still remain to some extent dependent upon lightness. The reason for this can best be appreciated by considering the shape of the Munsell colour solid (figure 6 a). In this, the achromatic series of greys between black and white form the axis from which the various hues radiate in increasing chroma. If a cross-section of this colour solid be taken by passing a plane through the achromatic axis, a typical section will be as shown in figure 6 b, which represents a cross-section through the yellow and the complementary purple-blue hues. The outer limits of this cross-section represent the colours of McAdam's limiting colour solid, after correction for surface-scattered light. It will be seen that in both halves of the section shown, the chroma first increases as the value decreases from white, but then reaches a maximum and

decreases again as the value approaches zero, or black. In other words, there can be no colour in either black or white, but only at intermediate values. It is important to note that different hues attain their maximum chroma at different values. Thus yellows can reach high chroma or colourfulness with small light absorption, but high chroma in purples can only be reached at low values, a phenomenon which is due to the variable sensitivity of the eye to the different colours of the spectrum.

In figure 6 the points representing dyeings of a typical real dye in gradually increasing strengths have also been plotted, and it will be seen that the maximum chroma attainable by a real dye is less than that of the limiting locus and, furthermore, is attained at a lower level of value. It is also clear that the number of chroma steps separating the real dye from the theoretical limit varies with the value level at which the comparison is made, so that the problem still remains of selecting for the comparison a value level which will be equally satisfactory for all hues. The best solution of this problem seems to lie in a comparison at the value level at which the ideal locus attains maximum chroma (A in figure 6), since as the real dye becomes more perfect so does its locus approach the limiting locus. The value level in question will of course differ for each hue.

It is perhaps desirable to digress a little here in order to consider a point raised by Mr. R. G. Horner in the discussion after the lecture. He suggested that in trying to define the efficiency of a dye in terms of a single number, we may be attempting the impossible. It may be that inefficiency can only be adequately described by a statement of both the chroma and value deficiencies of the real dye. This may well be the case, but it is felt that, from the dye manufacturer's point of view, a single number is always preferable to a more complex description, as it leads to no ambiguity as to which of two dyes is the better. Therefore, unless it can be shown conclusively that a single function is inadequate to describe the defects of dyes, there seems no reason to increase the complexity. Furthermore, by comparing the theoretical maximum chroma with the chroma of the real dye at the same level of value, instead of comparing the maximum chroma attainable by the real dye with the theoretical maximum, we are in effect introducing an empirical correction for the fact that the real maximum chroma is attained at a lower value level. Examination of figure 6*b* will make this point clearer, for it is proposed to measure the chroma deficiency indicated by the distance RM, rather than the difference between the maximum chromas ($AM - BR'$). From a rather superficial examination of the results obtained and comparison with visual observations on the brightness of the dyes examined, the proposed measurement appears to be fairly satisfactory, but it may well be found, on a more detailed scrutiny, that anomalies appear.

After due consideration it was decided to adopt the above proposal for the determination of inefficiency. Before this can be put into practice, however, it is necessary to know the actual shape of the Munsell colour solid, in order that the optimum value level may be selected for the comparison of a dye of any specific hue. The method of determining this shape was as follows. The tri-chromatic coefficients of all the Munsell colours of value 6 (for example) were plotted on a chromaticity diagram and isochroma lines were drawn through the points of equal chroma. The figures for the co-ordinates of the Munsell colours

were taken from the recent publication of the Optical Society of America, in which the spacings have been smoothed and extrapolated to the theoretical limits. On this diagram was superimposed the theoretical limit contour corrected for surface-scattered white light for colours of that particular lightness (30%) or Munsell value (Newhall *et al.*, 1943). The resulting diagram is shown in figure 8. Similar diagrams were prepared for value levels of 3, 4, 5, 7 and 9. Value levels below 3 were not studied, for a Munsell value of 2 corresponds to a lightness of only 3.4%, and since the patterns have a matt surface, at least 50% of the light reflected from them must consist of surface-scattered white light. The accuracy with which the theoretical limits may be defined is thus very low. In the present writer's opinion, Munsell patterns of value 1 (1.2% lightness) are unrealizable in practice, and the extrapolation of the chroma of patterns of this value to the spectrum locus is meaningless.

The resulting set of six diagrams of the type shown in figure 8 represents cross-sections of the Munsell colour solid. Cross-sections at right angles to these were then prepared by passing planes through the black-white axis at various

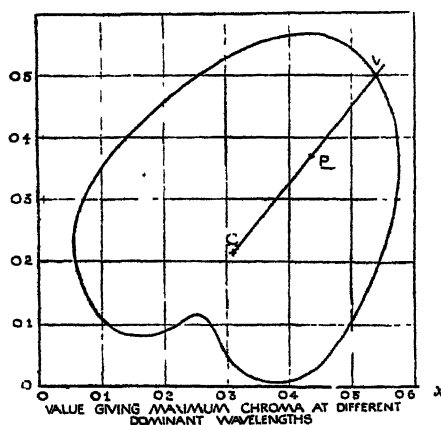


Figure 7.

hues and determining the limiting chroma at each value level. By plotting chroma against value, curves of the type of figure 6 were obtained, from which the value level yielding maximum chroma in that particular hue plane could be determined. Such cross-sections were cut at 26 different hues spaced fairly evenly round the colour circle, with a rather greater concentration around the ends of the spectrum locus, where abrupt changes may occur. The value levels yielding maximum chroma and the corresponding chroma are shown in table 3 but in practical use it is more convenient to express the results in graphical form, shown in figure 7. This diagram is constructed on the standard chromaticity diagram, and is so contrived that if a straight line is drawn through C to cut the heart-shaped locus at V, the distance CV indicates the value at which the ideal dye having the hue of points on CV attains maximum chroma.

The whole procedure involved in determining the inefficiency of any real dye will now be reviewed briefly, in order to clarify the position and indicate the practical steps which are involved. A dyeing is first prepared and measured in

the Donaldson colorimeter or in any other suitable instrument. The trichromatic co-ordinates are plotted in figure 7 at P and the line CP drawn to cut the locus at V. The distance CV is measured in units of value (1 unit of value = 0.04 units of x), and this indicates the value at which the ideal colour of the same hue at P attains its maximum chroma. If the value of the actual pattern does not coincide with the value so determined, it must be varied by increasing or decreasing the amount of dye applied to the cloth until it is of the correct value. Fortunately,

Table 3. Value levels yielding maximum chroma on cross-section through modified colour solid

Cross-section through C point and point in column below		Value level yielding maximum chroma	Maximum chroma
x	y		
0.310	0.700	7.7	20.5
0.400	0.700	9.1	18.0
0.500	0.700	9.5	17.0
0.600	0.700	9.1	17.0
0.700	0.594	7.8	16.5
0.700	0.480	7.0	17.0
0.700	0.400	6.3	18.5
0.700	0.360	6.0	19.0
0.700	0.316	5.6	19.5
0.700	0.200	5.4	20.5
0.600	0.104	5.4	22.0
0.500	0.000	5.4	26.0
0.366	0.000	5.0	30.0
0.310	0.000	4.5	32.0
0.198	0.000	3.0	[35.0]
0.122	0.000	3.0	[30.0]
0.000	0.000	4.8	19.0
0.000	0.145	5.8	16.0
0.000	0.226	6.0	17.0
0.000	0.316	6.1	19.5
0.000	0.400	6.1	21.5
0.000	0.500	6.2	24.0
0.000	0.600	6.2	25.0
0.000	0.800	6.5	26.5
0.100	0.800	6.6	27.0
0.200	0.900	7.2	24.5

the hue does not usually change greatly with depth of shade, for if it did, it would be necessary to proceed by successive approximation, redetermining the optimum value level for each new pattern until coincidence was obtained. A pattern of the correct value having been obtained, its trichromatic co-ordinates are determined and plotted on a cross-section of the colour solid at the same value. This cross-section is mapped out with the appropriate isochroma contours as shown in the example in figure 8, and by drawing a line from C through P to cut the limiting locus at L, the number of steps of chroma improvement possible is immediately evident.

Proceeding in this way, the inefficiency of a number of dyes has been determined. The results are given in detail in table 4, which shows the percentage depth of shade needed to produce the required value for the particular hue of each dye, the trichromatic co-ordinates of the corresponding pattern and the actual lightness (corrected for the absorption of the substrate), the dominant wavelength, value, chroma, limiting chroma, and chroma improvement possible. The dyes are arranged in the table in order of hue. It should be noted that the dyeings here presented are not necessarily of the same depth of shade as those used in the earlier investigations on colorimetric and excitation efficiency.

From the rather limited number of results available, it appears that reds and yellows are highly efficient, blue-greens and bluish-reds less so, and purples and greens very inefficient. It seems unlikely that this variation can again be explained by any systematic error in the method of expression. Examination of typical absorption curves for the six primary colours (additive and subtractive)

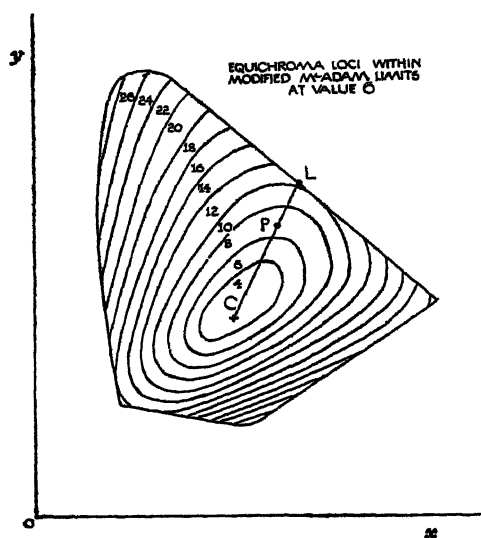


Figure 8.

shows that only in green and purple are there two absorption band edges which are both in the visible spectrum. For the production of the other four colours it is necessary only to have one band edge lying in the visible region. Since it is the sloping character of this band edge as compared with the vertical band edge of the ideal colours which produces inefficiency, it is obvious that greens and purples are almost certain to be the least efficient of colours, and this is confirmed by the present measurements. There is no obvious reason why blue-green dyes should be less efficient than yellows, but in actual fact they are less efficient, and this is probably due to the presence in all the dyes examined of a small secondary absorption in the far blue.

The case of Rhodamine BS is exceptional and interesting. It has a very high efficiency by any criteria, and, indeed, at certain lightnesses its excitation or colorimetric efficiency exceeds 100%. In other words, the points representing the pattern lie outside the colour solid. This result was regarded with some

suspicion at first, but repeated measurements have confirmed its truth. Consideration of the known fluorescent properties of this dye indicated that such a behaviour might result if the dye absorbs visible radiation and then re-emits part or all of this as visible radiation of longer wave-length. A simple experiment sufficed to show that this was the true explanation. The standard magnesium oxide screen, a piece of white wool and a piece of wool dyed with Rhodamine BS were placed in the Donaldson colorimeter, illuminated with white light, and measured with a Wratten trichromatic red filter interposed between the eye and the pattern. The relative lightness of the patterns measured in this way was: magnesium oxide 98%, white wool 75%, Rhodamine dyeing 117%. Thus the Rhodamine BS is absorbing green light and re-emitting it as red, and in consequence the patterns can lie outside the McAdam colour solid, which is based on the assumption that surface colour is produced by light absorption only.

§ 6 CONCLUSIONS

From the above results it is considered that the chroma deficiency of a dye at the optimum value level of the appropriate hue provides a satisfactory measurement of the efficiency of the dye and the possibilities of visual improvement in brightness which remain. The results obtained by this method, on a number of dyes of differing hues, seem to show that those colours which have only one absorption band edge in the visible spectrum are more efficient than those with two, as might be expected.

From a more practical aspect it is considered that this method of measurement provides a sound basis for the quantitative comparison of the brightness of dyes of all hues, although it will be most precise and possibly most useful when the dyes compared are of similar hue. The results have also shown that in certain colours commercial dyes leave little room for improvement.

§ 7 ACKNOWLEDGMENT

The author wishes to express his thanks to Imperial Chemical Industries (Dyestuffs Division) for permission to publish this paper.

REFERENCES

- BRECKENRIDGE, F C and SCHRAUB, W R, 1939 *J Opt Soc Amer* 29, 371
JUDD, D B., 1935 *Ibid* 25, 24.
KELLY, K L., GIBSON, K S and NICKERSON, D, 1943 *Ibid* 33, 355
MCADAM, D L, 1935 *Ibid* 25, 361
NEWHALL, S. M., NICKERSON, D and JUDD, D B, 1943 *Ibid* 33, 385.
WHITE, G S J, VICKERSTAFF, T and WATERS, E., 1943 *Proc Phys Soc* 55, 1

RESONANCE IN PRECESSIONAL STATES OF DIATOMIC MOLECULES

By H. G. HOWELL,
University College, Southampton
(Now at Technical College, Bradford)

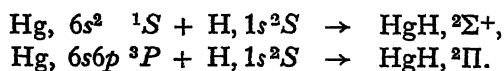
MS received 27 May 1944

ABSTRACT. Examples of electronic states described by van Vleck as "pure precession" are found in the Hg, Cd and Zn halide spectra and also with InO and GaO. These are shown to be in close resonance and probably have abnormally large Λ -type doubling. An interpretation of the spectra of InO and GaO is given and a potential-energy diagram is presented for these molecules and also for TiO.

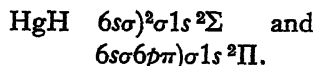
§1. INTRODUCTION

IN van Vleck's theoretical treatment of Λ -type and spin doubling in electronic states of diatomic molecules, it is shown that simple relations can be derived for the doubling constants for Σ and Π states if they are related to one another in a manner described by him as "pure precession". Such states act as though they have identical electron configurations, except that one has $\Lambda=0$ in the Σ state and $\Lambda=1$ in the corresponding Π state, whilst each retains the same atomic L vector, which remains quantized in the molecule whilst precessing round the electric axis. If there is only one valence electron outside closed shells, then λ and l here replace Λ and L . In addition, even the principal quantum number n of an atom may retain its significance on entry into a molecule.

Mulliken and Christy extended this work and applied the results to a number of molecules for which doubling data were available, and were able to verify van Vleck's work with the following hydrides: BeH, MgH, CaH, ZnH, CdH, HgH, CH, SiH and OH, and also with the homonuclear molecules He₂, Li₂ and Na₂. With the hydrides the results were interpreted on the assumption that the molecule is similar to the united atom, with the electrons, except that of H, retaining their atomic quantum numbers. For example, it is known that the normal $^2\Sigma$ and excited $^2\Pi$ states are related to the separated atoms as follows:



If the H electron in the molecule were still essentially a $1s$ electron, it would be possible to write



The electron configuration of both states would be the same except for the replacement of the $6s\sigma$ by the $6p\pi$. Now these orbits differ in l and hence cannot be related precessionally, whereas evidence from the Λ -type doubling shows

that the spectrum derives from two states which are so related. Hence it must be supposed that the $H\ 1s$ electron becomes a $6p\sigma$ in the molecule, thus giving $6s\sigma)^2 6p\sigma, {}^2\Sigma$ and $6s\sigma)^2 6p\pi, {}^2\Pi$ for the two states being similar to the $6s^2 6p$ configuration of the united atom Tl. This example illustrates the importance of the Λ -type doubling in deciding which of possible configurations is the correct one.

With He_2 , Li_2 and Na_2 the results were similarly interpreted in terms of the quantum numbers of the separated atoms.

The magnitude of the Λ -type and spin doubling is dependent upon the separation of the Σ and Π states, $\nu_{\pi\Sigma}$ being larger the smaller the separation, i.e. the closer the two precessional states the greater the resonance shown in the fine structure. The largest Λ -type doubling coefficient discussed by Mulliken is that of CaH , where $\nu_{\pi\Sigma}$ is 1300 cm^{-1} .

The objects of this paper are (a) to point out the existence of even closer resonance between precessional states in the Zn , Cd and Hg halides, and (b) to interpret the spectra of the oxides of In , Ga and Tl in terms of similar close resonating precessional states.

§ 2. RESONANCE IN Zn , Cd AND Hg HALIDES

In a recent paper the author showed that the u.v. system of bands for the molecule HgF were due to a ${}^2\Pi - {}^2\Sigma$ transition (Howell, 1943). It was also found that there existed a close relation between the splitting of the ${}^2\Pi$ state and the coupling factor a of the $6s6p\ ^3P$ state of Hg , a relation which could only be interpreted as indicating that the $Hg\ 6p$ electron retained its l value in the molecule, becoming a $p\pi$. Further, the close proximity of the system to the Hg ground line $*\ ^3P_1 - {}^1S$ meant that the principal quantum number n was also retained in the molecule. In short, the transition was the molecular analogue of the atomic transition $6s$ to $6p$, and can be described as occurring between atomic orbitals $6s\sigma$ and $6p\pi$. The electron configurations were therefore represented as $\pi^4 6s\sigma 6p\pi\ ^2\Pi - \pi^4 6s\sigma\ ^2\Sigma$.

Probably a better representation would be $6s\sigma 6p\pi)\pi^4\sigma\ ^2\Pi \leftarrow 6s\sigma)^2\pi^4\sigma\ ^2\Sigma$, which implies that the Hg atom largely retains its quantum numbers in the molecule and that the molecular binding is due to the $\pi^4\sigma$ coming from the halogen.

Replacing $4s$ and $5s$ for the above $6s$ gives possible ZnF and CdF configurations, which have now been identified, as both these molecules give the same type of atomic-like transitions. The same band systems were also sought for and found among unanalysed data in nearly all the other halides of these atoms.

Now the ${}^2\Pi$ and ${}^2\Sigma$ states thus established are not related precessionally because they derive from different l values, but the state $6s\sigma 6p\pi)\pi^4\sigma\ ^2\Pi$ should have such a precessional state $6s\sigma 6p\sigma)\pi^4\sigma\ ^2\Sigma$ related to it. Attention was called to the fact that in most of the halides investigated, the ${}^2\Pi - {}^2\Sigma$ system was mixed up with a fragmentary one having the same vibrational constants, this confusion being one of the difficulties met with in the analysis. In the case of ZnF , Rochester (1939) definitely attributed these additional bands to a ${}^2\Sigma - {}^2\Sigma$ transition, and there seems no doubt that his allocation is correct, in which case

* It is proposed to use the term *ground line* throughout this work in preference to "resonance line" to avoid confusion with the quantum resonance which is discussed in the text.

the upper ${}^2\Sigma$ state must be the $4s\sigma 4p\sigma\pi^4\sigma^2\Sigma$. The reason why such a definite statement can be made is the occurrence of the ${}^2\Sigma - {}^2\Sigma$ always in the same region of the atomic ground line as the ${}^2\Pi - {}^2\Sigma$, showing that both systems are intimately related to the atomic ${}^3P_1 - {}^1S$ line.

The closeness of these two precessional levels ${}^2\Sigma$ and ${}^2\Pi$ to each other means that very strong resonance can be expected between them. Unfortunately the data on the ${}^2\Sigma$ state are so incomplete—in some cases its existence can only be inferred—that the difference in energy $\nu_{\pi,\Sigma}$ between the two states cannot be given. In ZnF, ${}^2\Sigma$ lies slightly above ${}^2\Pi$, whilst in the Hg halides it lies between the two ${}^2\Pi$ components, i.e. the separation is less than the $\Lambda - \Sigma$ coupling. In most cases it is certain that the resonance is closer than in any molecule hitherto studied. In such cases the L vector must be almost completely uncoupled from the nuclear axis. Now van Vleck's theoretical relations are only applicable to cases in which $\nu_{\pi,\Sigma}$ is large compared with the $\Lambda - \Sigma$ coupling coefficient A , and in this case this is far from being fulfilled. It can be anticipated, however, that for these precessional states the Λ -type and spin doubling will be very large and probably greater than the rotational levels themselves.

Certainly interesting results can be expected from a rotational analysis of any of these systems, but the prospects of such an analysis being made are remote because of the close structure of these heavy molecules, e.g. the ${}^2\Pi_{1/2} - {}^2\Sigma$ system of HgF is represented mainly by Q branches, each of which is condensed into quite a narrow line, but it may be possible to resolve the ZnF structure.

It is interesting to note that the fading out of the P and R branches in the above system may be a consequence of the close resonance, as well, maybe, as the diffuse appearance of the ${}^2\Sigma - {}^2\Sigma$ system in other cases.

§ 3. APPLICATION TO THE SPECTRA OF Ga, In AND Tl OXIDES

Isoelectronic with the halides already discussed are the molecules GaO, InO and TlO, and it is reasonable to expect a close similarity of electron configuration for these molecules. From Ga, $6s^2 6p^2 P + O, 2p^4 {}^3P$ the configuration GaO, $\sigma^2 \sigma^2 \pi^3, {}^2\Pi$, could result using molecular orbitals, but examination of the known spectra shows that a construction similar to the Hg halides must be assumed.

(a) *GaO*. This is found in a Ga arc in air and has been studied most recently by Guernsey (1934). The bands occur near the Ga ground lines ${}^2S - {}^2P$ (2P separation 826 cm^{-1}), most having double heads with an average separation of 2.6 cm^{-1} .

These are described as R heads of a ${}^2\Sigma - {}^2\Sigma$ system and represented by

$$r = \frac{25706.43}{9.04} \left\} + (763.63u' - 3.89u'^2) - (767.69u'' - 6.34u''^2).$$

Evidence is mentioned of a second system lying in the $4300\text{--}4900\text{ \AA}$. region, the bands being faint and almost entirely obscured by a continuum attributed to Ga₂, but no data are given for this system.

(b) *InO*. This spectrum has been investigated by Watson and Shambon (1936), who produced it in exactly the same way as for GaO. Band heads are observed from 3847 to 4763 \AA . the strong ones being confined to the region of

the In ground lines $^2S-^2P$ (2P separation 2122 cm. $^{-1}$) and are considered to be due to either a $^3\Pi-^2\Sigma$ or a $^2\Sigma-^3\Pi$ system. Watson and Shambon favour $^3\Pi-^2\Sigma$ in view of $^2\Sigma$ being apparently the ground state of GaO. Their quantum formula for this system is

$$\nu = \frac{23595 \cdot 10}{23033 \cdot 10} \left\{ + (626 \cdot 66u' - 3 \cdot 40u'^2) - (703 \cdot 09u'' - 3 \cdot 71u''^2 - 0 \cdot 285u''^3), \right.$$

the cubic term being necessary to cover the rapid convergence of the lower state.

There is evidence of another system interlaced within this doublet system.

§ 4. INTERPRETATION

The account of the spectra of these molecules in terms of electron configurations, which will follow, is based upon the following considerations:—

(1) The spectra, produced in the same way, belong to very similar molecules, and should, therefore, be due to the same transition, i.e. they should be both $^3\Sigma-^2\Sigma$ or both $^3\Pi-^2\Sigma$ or $^2\Sigma-^3\Pi$. That this is apparently not so demands an explanation.

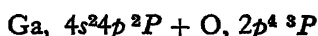
(2) Both systems are corresponding ones in that they both occur in the region of the corresponding atomic ground lines.

(3) The vibrational frequencies are very similar in upper and lower states of both molecules.

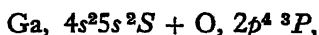
(4) In each case there are signs of the existence of another system in the same region.

The occurrence of both spectra near the atomic ground lines suggests that, as in the Hg halides, the transition is between atomic orbitals related to the 2S and 2P orbits. This is borne out by the near equality of the vibrational frequencies, which is a characteristic feature of such transitions.

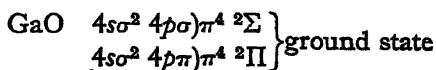
Now the ground state must derive from



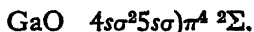
and the excited state



so that the molecular electron configurations must be

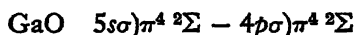


and the excited state



The two possible configurations for the ground state are linked precessionally, an interesting difference from the Hg halides, where the precession states are excited. The foregoing applies also to InO by replacing $4p$ and $5s$ by $5p$ and $6s$.

The observed band systems can now be identified as



and



i.e. the InO system is not $^2\Pi-^2\Sigma$ as Watson and Shambon believed. Which of the two possible ground states is the lower is not certain, as this depends upon a more complete study of the unassigned bands mentioned in both investigations. With GaO it is certainly probable that $^2\Sigma$ is the ground state with, possibly,

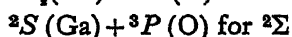
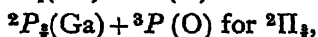
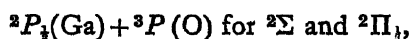
the 4900 Å. system, as the ${}^2\Sigma - {}^2\Pi$; but with InO it looks as though the resonance is so close that both states have an equal claim to be considered the ground state.

It is not at all certain that Watson and Shambon's identification of the 560 cm^{-1} interval as the ${}^2\Pi$ separation is correct, for this doubling is not of the expected order. This should be due entirely to the $5p\pi$ electron, and is related to the coupling factor of the $5p$ electron of In. The magnitude of the expected ${}^2\Pi$ splitting turns out to be 1475 cm^{-1} , and in view of the strongly atom-like nature of the transition it would be unusual if the correspondence did not extend to the magnitude of the coupling.

The two levels which Watson and Shambon have linked together as ${}^2\Pi$ may possibly turn out to be a ${}^2\Sigma$ and one component of the ${}^2\Pi$. This ${}^2\Sigma$ will have vibrational constants almost identical with the ${}^2\Pi$, and so would fit into a doublet scheme without difficulty. It is evident that there is a need for a further experimental study of this molecule.

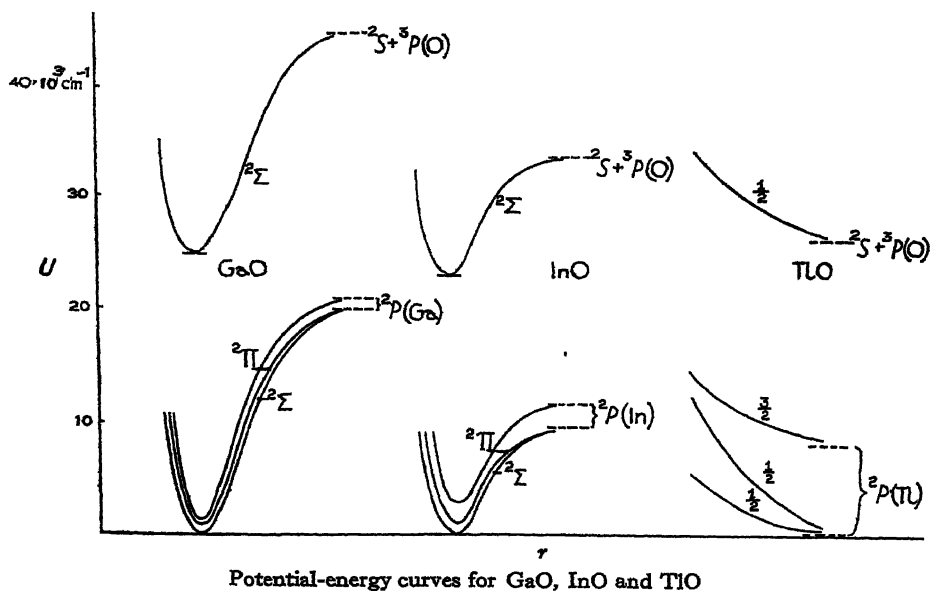
§ 5. DISSOCIATION ENERGIES AND POTENTIAL-ENERGY CURVES OF GaO, InO AND TlO

With these molecules there is no doubt as to the nature of the dissociation products, viz.:



and also for InO.

This fact, together with the rapid convergence of the ground state in the



case of InO, renders it possible to calculate fairly reliable values of both D' and D'' for this molecule. This gives $D'' = 9000$ and $D' = 11,000\text{ cm}^{-1}$. Using direct extrapolation for the GaO ground state and assuming that a similar tendency exists for a slightly more rapid convergence than is given by the formula, a value of $D'' = 20,000\text{ cm}^{-1}$ is obtained. This gives $D' = 19,000\text{ cm}^{-1}$, which

shows that in both molecules both states have practically the same D values, and hence almost identical potential-energy curves. These are sketched in the figure, assuming $^2\Sigma$ to be the lowest in each case.

Watson and Shambon report that they found no bands in the Tl arc in air in the neighbourhood of the Tl ground lines, and concluded that one or both of the corresponding TlO levels were unstable. That this is probably the case can be inferred from the big reduction in stability in going from GaO to InO (20,000 to 9000 cm^{-1}). Since there is a large increase in the weight of the TlO molecule, there seems little chance that any of the levels will have even a flat minimum in their potential-energy curves. The probable course of these is sketched in the figure, assuming that a case-c coupling tendency is marked with this heavy molecule, and so each level is represented by its Ω value.

REFERENCES

- GUERNSEY, M., 1934. *Phys Rev.* **46**, 114.
 HOWELL, H. G., 1943. *Proc. Roy. Soc. A*, **182**, 95
 MULLIKEN, R. S. and CHRISTY, A., 1931. *Phys Rev* **38**, 87
 ROCHESTER, G. D., 1939. *Z Phys.* **114**, 495
 VAN VLECK, J. H., 1929. *Phys Rev* **33**, 467.
 WATSON, W. W. and SHAMBON, A., 1936. *Phys Rev* **50**, 607.

THE SPECTRA OF TIN AND LEAD HYDRIDES

By H. G. HOWELL,

University College, Southampton

(Now at Technical College, Bradford)

MS received 23 May 1944

ABSTRACT. It is shown that the 2II ground-state doubling of SnH is due to a $p\pi$ electron, and its magnitude can be predicted from the 3P width of the Sn ground state. The same doublet separation is found among the red bands of SnH, and a vibrational analysis is presented for this system, which is considered to be a $^2\Sigma-^2II$. It is assumed that PbH will also have a 2II ground state with an estimated separation of 8200 cm^{-1} and which is modified by case-c coupling. Although this interval is actually found among the PbH bands, its occurrence is not taken as proof of the existence of the 2II ground state, as it appears to be more likely the separation of two excited states. The red system analysed by Watson and Simon and described by them as $^2\Sigma-^2\Sigma$ is here interpreted as a $\frac{1}{2}(^2\Sigma)-\frac{1}{2}(^2II)$ system analogous to the SnH $^2\Sigma-^2II$ system, and an isolated band at 3815 Å. is considered to be the 0,0 band of $\frac{3}{2}(^2\Delta)-\frac{1}{2}(^2II)$. It is explained why the main band system has the appearance of a $^2\Sigma-^2\Sigma$ system and why the spectra of SnH and PbH are apparently so dissimilar.

The occurrence of two GeH systems is predicted. A similarity between the hydrides and halides of this carbon group is described and a number of 2II separations for the ground states of these molecules is predicted.

§1. INTRODUCTION

THE spectrum of lead hydride, PbH, has been investigated by Watson (1938) and by Watson and Simon (1940), using a high-pressure arc in hydrogen. They found bands extending from 5000 Å. to 9100 Å., apparently belonging to one system which, after a rotational analysis had been made, was

considered to be a ${}^2\Sigma - {}^2\Sigma$. This conclusion was reached even though some of the bands had branches whose intensities seemed out of order for such a transition and which favoured rather a ${}^2\Pi - {}^2\Pi$ transition. This latter possibility was finally dismissed, owing to the absence of the expected sub-bands and of certain combination differences.

The same workers (1939) have also studied tin hydride, SnH, using the same method of excitation, and they identified bands at 4054 and 4447 as belonging to a ${}^2\Delta - {}^2\Pi$ system with a normal ${}^2\Pi$ state ($A = 2183 \text{ cm.}^{-1}$) and an inverted ${}^2\Delta$ with a very small splitting ($A = -1.75 \text{ cm.}^{-1}$). In addition, weaker bands were found in the red, but a satisfactory analysis of these could not be made. The only bands listed were at 6745, 6892, 7032 with a pile-up of lines at 6931 and also at 6095 and 6214 Å. A certain number of ${}^2\Pi$ -state combination differences were found among lines around 6931, and Watson and Simon consider that these bands probably belong to a ${}^2\Sigma - {}^2\Pi$ transition, as they are quite different from the bands of the red ${}^2\Sigma - {}^2\Sigma$ system of PbH.

The surprising feature of this comparison of these two molecules is the complete lack of any similarity between their spectra. An attempt to obtain the analogous ${}^2\Delta - {}^2\Pi$ system of PbH was actually made by Watson and Simon, but they found only a weak band at 3815 Å. which was "not accompanied by another at any plausible interval for a ${}^2\Pi$ PbH state". This lack of correspondence is surprising in view of the close resemblance of tin and lead, which in the author's opinion is sufficient to guarantee a similarity of electron configuration in corresponding molecules. The common mode of excitation and the need for high pressures in both cases suggest that the same type of electronic levels is involved. It is particularly difficult to interpret the ${}^2\Sigma - {}^2\Sigma$ system of PbH in terms of electron configurations. Watson and Simon accounted for these levels by assuming that they are formed from Pb, $6p^2 + \text{H}, 1s$, the latter being promoted to a $7s$ orbit to give $6p\sigma^2 7s\sigma^2 \Sigma$ and $6p\pi^2 7s\sigma^2 \Sigma$. Such a transition involves a two-electron jump, and one cannot but feel uneasy over such an explanation, particularly in view of the absence of any normal transition. The levels normally expected would be $6p\sigma^2 6p\pi^2 \Pi$ and $6p\sigma 6p\pi^2 \Delta$, ${}^2\Sigma^+$, ${}^2\Sigma^-$, all of which are found in CH.

The difference in spectral types in SnH and PbH therefore presents a problem which it is hoped this paper will solve. It will be shown that there are actually points of resemblance between them, and that the observations of Watson and Simon can be included within a consistent scheme. In particular, an alternative interpretation of the red ${}^2\Sigma - {}^2\Sigma$ bands will be given and an analysis of the red SnH bands will be suggested.

§ 2. COMPARISON WITH SIMILAR HYDRIDE MOLECULES

Much can often be learnt and predicted about the various band systems of a molecule by comparing them with the well-established systems of a similar molecule. By *similar molecules* are meant those in which one atom is common, whilst the others belong to the same group of the Periodic Table. This is different from the "similarity" of iso-electronic molecules such as N_2 and CO, and it is proposed to differentiate by calling the similar molecules here discussed *group molecules*. Thus, in the present case the molecules SnH and PbH will

be compared with their group molecules CH, SiH and GeH. However, as yet nothing is known about GeH, and so we are restricted to CH and SiH. CH has a $^2\Pi$ ground state with a doublet separation of 28.5 cm^{-1} with $^2\Delta$, $^2\Sigma^+$ and $^2\Sigma^-$ in that energy order. SiH also has a $^2\Pi$ ground state with the larger separation of 124 cm^{-1} , and the only known excited state is the corresponding $^2\Delta$. Thus, assuming a close similarity between group molecules, the $^2\Pi$ state of SnH can be safely identified as the ground state and a similar level with a still larger separation

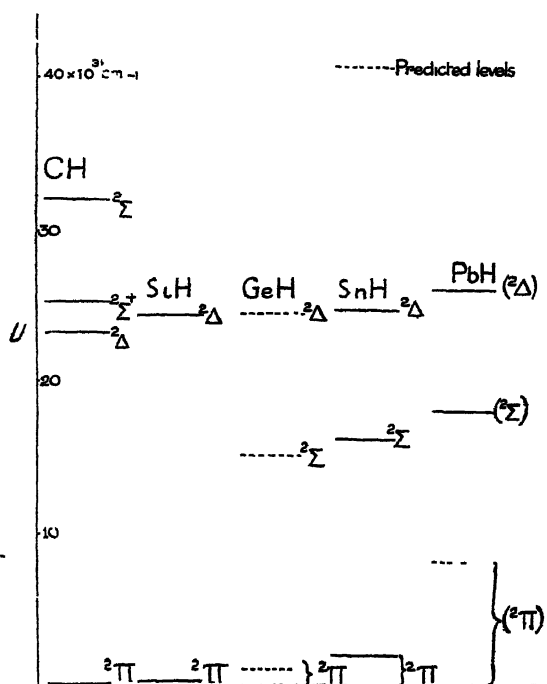


Figure 1. Energy levels of CH-type group molecules

can be expected to be the ground state of PbH. In figure 1 are collected all the known states of these molecules together with predicted states which will be described in due course. An obvious line of enquiry lies in the investigation of the $^2\Pi$ separations with the aim of predicting that expected for a PbH state. Another is to explore the possibility of other states competing with the $^2\Pi$ for the ground state. There seems no likelihood of this in CH or SiH, nor is any such state predicted from a study of the electron configurations.

§ 3. RELATION BETWEEN ATOMIC AND MOLECULAR COUPLING FACTOR

The author showed that with HgF and its group molecules the atomic coupling factor α of the outer p electrons was related to the coupling coefficient A of the molecule which determines the $^2\Pi$ separation (1943). This relation indicated that the atomic p electron became a $p\pi$ in the molecule, and it enabled the $^2\Pi$ separation to be determined for a number of these molecules whose spectra had remained unanalysed. By using the predicted interval as base, it was found possible to present satisfactory analyses. In the present case

the doubling in the $^2\Pi$ state is due to a single π electron, and it is possible that a similar relation holds. The doublet separation should be equal to the atomic coupling coefficient a of the corresponding p electron. Table 1 gives these a values and the $^2\Pi$ interval $\Delta\nu$ where known.

Table 1. Comparison of a (from 3P) with $^2\Pi$ interval

	CH	SiH	GeH	SnH	PbH
a (calc.)	28	149	940	2285	$\left\{ \begin{array}{l} 7800^* \\ 8200^\dagger \end{array} \right.$
$^2\Pi$ (observed)	29	124	—	2183	

* From a of Pb^+ (3P).

† Independent estimate.

Neglecting PbH for the moment, it can be seen that the agreement for CH, SiH and SnH is sufficient to confirm the view that here too the p electron definitely becomes a $p\pi$ electron. Further, the agreement for SnH seems a justification for expecting an equally good correspondence for PbH, and encourages the view that the $^2\Pi$ interval for this molecule can be predicted. Now the value of a cannot be determined directly from the $6p^2\ ^3P$ width of the Pb ground state as in the other cases, as the separation of the 3P components shows a big departure from normal L-S coupling. It is, however, possible to estimate the value from $6p\ ^2P$ of Pb^+ , as comparison of the a values derived from Si and Si^+ and Sn and Sn^+ reveals that the ionized atom gives a value 28 % and 23 % too high for Si and Sn respectively, and so, assuming that a from Pb^+ will be 20 % too high, a value of 7820 $cm.^{-1}$ is obtained. That this is probably on the low side is indicated by an independent method to be described later, which gives a value of 8230 $cm.^{-1}$ for the doublet width of $^2\Pi$.

The observation that with Pb a radical change of coupling has taken place is significant, for it means that if a change of term type does occur with any member of this molecular group, it should take place with PbH, for any change in the atomic constitution should be reflected in the molecule.

It begins to appear, therefore, that differences are to be expected in the PbH spectrum. This does not mean that the problem has been removed, for it is still necessary to interpret the apparent $^2\Sigma$, $^2\Sigma$ states and to explain how they have changed from the expected $^2\Delta$ and $^2\Pi$. It is probable that such a change has come about by alterations in the type of coupling, but it should always be possible to trace this change and to correlate the observed levels with those expected for normal coupling. Before this is done it will be found advantageous to clear up the explanation of the red bands of SnH.

§ 4. THE RED SYSTEM OF SnH

By the number of observed branches and the occurrence of some combination differences among a few of the red bands of SnH, a $^2\Sigma - ^2\Pi$ transition was suspected by Watson and Simon. Assuming that $^2\Pi$ is the ground state, it follows that, unless only one component of the doublet is present, there must occur the ground-state interval of 2183 $cm.^{-1}$ among these bands. Of the five listed, two do show such a difference, viz. 6095 Å. and 7030 Å., which have a

separation of 2181 cm.^{-1} , which seems too close to the expected value to be a coincidence. Assuming these to be the 0,0 bands of the system, they can be arranged in the following vibrational scheme:—

Table 2. Vibrational scheme for red bands of SnH

$v'' \backslash v'$	0		1
0	16402 2181 14221 (1580)		
1	14822	(1266)	16088 (1582)
2			14506

The missing members of the ${}^2\Sigma - {}^2\Pi$ system should lie to the red of the last measured band, 7030 \AA ., and Watson and Simon mention that more bands do lie to the red of this. There should also be members of the $\Delta v = 1$ sequence found on over-exposed plates. If this arrangement is accepted, then ω_e'' and ω_e' are approximately 1600 and 1300 cm.^{-1} respectively. Although the value of ω_e'' cannot be checked from the ${}^2\Delta - {}^2\Pi$ system (as this is limited to the 0,0 bands) it is of the right order, as it should be between the value of 2000 for SiH and 1560 cm.^{-1} for PbH, and necessarily nearer to the latter, since another value, that of GeH, must intervene. This assumes that the value quoted refers to the ground state of PbH. To complete this brief account of the SnH spectrum, it remains to note that according to the probable electron configuration of the ${}^2\Delta$ state, it should be normal, and so it is probable that the small negative value observed for A is due to the interaction of the s of other electrons with the l of the valence electron.

§ 5. INTERPRETATION OF THE PbH SPECTRUM

If it be assumed that the ground state of PbH is ${}^2\Pi$, then a separation of the order of 8200 cm.^{-1} ought to be found in the spectrum. That such a difference does occur has been overlooked by Watson and Simon. This is between the system origin of the red bands and the solitary band at 3815 \AA ., and amounts to 8187 cm.^{-1} . This is so close to the expected value that it is tempting to conclude that the existence of the supposed ${}^2\Pi$ ground state is fully established, and that PbH is quite similar to its group molecules. If this is so, it remains to decide whether the upper state is ${}^2\Sigma$ or the expected ${}^2\Delta$, and then to enquire why the lower state behaves like a ${}^2\Sigma$.

A case can be made out for ${}^2\Sigma$ or ${}^2\Delta$, or even for both together, being the upper state of the red bands, on the hypothesis that PbH has a partial case-c coupling.

For pure case-c, the components ${}^2\Delta_1$ and ${}^2\Delta_3$ will give rise to the levels $\Omega = \frac{5}{2}$ and $\frac{3}{2}$, which can combine with the case-c ground levels derived from ${}^2\Pi$. This should result in an increase in the number of systems as compared with SnH. If, however, the coupling is only partial case-c, then those transitions governed by $\Delta\Omega = 0$ will be favoured and others may be too weak for observation. Thus transitions $\frac{1}{2}({}^2\Delta) - \frac{3}{2}({}^2\Pi)$, $\frac{5}{2}({}^2\Delta) - \frac{3}{2}({}^2\Pi)$ and $\frac{3}{2}({}^2\Delta) - \frac{1}{2}({}^2\Pi)$ would be expected, among others, for pure case-c, but for partial case-c the first would be the most intense. Tentatively identifying the red system as $\frac{3}{2} - \frac{3}{2}$, the weak band at 3815 Å. can be described as $\frac{3}{2} - \frac{1}{2}$, whilst $\frac{5}{2} - \frac{3}{2}$ can be assumed to be similarly weak and covered by the stronger red system. Now a case-c $\frac{3}{2} - \frac{3}{2}$ system will have a structure similar to a ${}^2\Pi_3 - {}^2\Pi_3$, which, apart from the lowest J values, will resemble a ${}^2\Sigma - {}^2\Sigma$ transition. Such a view would account for the observed ${}^2\Sigma - {}^2\Sigma$ system, with the exception of one important detail—the so-called spin doubling of the lower state, which was found to be proportional to $K + \frac{1}{2}$. This must now be accounted for as Ω -type doubling and should be proportional to $(J + \frac{1}{2})^2$. Another difficulty is found on a closer comparison with SnH. The main system is limited to the 0,0 bands, and the 3815 PbH band $\frac{3}{2}({}^2\Delta) - \frac{1}{2}({}^2\Pi)$ is similarly isolated, which makes it hard to understand why the other component is represented by such a comparatively prolific number of bands. Since the red SnH system ${}^2\Sigma - {}^2\Pi$ has more bands than the ${}^2\Delta - {}^2\Pi$ (due probably to the shift of the upper potential-energy curve to larger r values) an alternative interpretation of the PbH spectrum is suggested. This is to associate the red system with a corresponding ${}^2\Sigma$ level (case-c $\frac{1}{2}$). The transition $\frac{1}{2}({}^2\Sigma) - \frac{1}{2}({}^2\Pi)$ will be similar to a ${}^2\Pi_1 - {}^2\Pi_1$ transition, which again resembles a ${}^2\Sigma - {}^2\Sigma$, and furthermore, the Ω -type doubling will now be proportional to $J + \frac{1}{2}$, which satisfies the observations of Watson and Simon. It should be recalled at this stage that the branch intensities of certain bands were in keeping with a possible ${}^2\Pi - {}^2\Pi$ transition. Further support for this new interpretation is obtained from figure 1, where it is seen that the $\frac{5}{2}$, $\frac{3}{2}({}^2\Delta)$ level in PbH is slightly higher than the corresponding level in SnH (actually 1606 cm.⁻¹). This difference is probably related to the difference in the atomic ionization potentials: Pb, 7.38 volts—Sn, 7.30 volts = 650 cm.⁻¹. Assuming a similar difference in the ${}^2\Sigma$ type levels, that for PbH would be expected to occur at 18060 cm.⁻¹, which is practically the value of ν_e for the red system. In this case the observed 8200 cm.⁻¹ separation is not due to the ${}^2\Pi$ at all, but represents the difference between $\frac{1}{2}({}^2\Sigma)$ and $\frac{3}{2}({}^2\Delta)$. This coincidence does not invalidate the previously developed argument for PbH having a case-c equivalent ${}^2\Pi$ ground state with a doubling of the order given, as the evidence given in table 1 is considered conclusive. On this view, then, the main red system is a $\frac{1}{2}({}^2\Sigma) - \frac{1}{2}({}^2\Pi)$ transition and the 3815 cm.⁻¹ band is $\frac{3}{2}({}^2\Delta) - \frac{1}{2}({}^2\Pi)$. The $\frac{3}{2}({}^2\Delta) - \frac{3}{2}({}^2\Pi)$ component ought to lie in exactly the same region as the red system because of the ground-state separation of about 8000 cm.⁻¹. It is rather tempting to consider that there might be two systems here, and Watson and Simon actually thought this a possibility because of the smallness of ω_e' . Its value is only half of what it should be according to the Mecke rule $\omega/B = \text{constant}$, which could only be satisfied by assuming the

existence of two separate systems which take the alternate v' values of the system as actually analysed by Watson and Simon. They decided against this, and the author agrees with their decision on the grounds that the similarity with SnH indicates that the system should be represented by only a weak and single band which is probably masked by the main system. The exact value of the ground-state separation is then left undetermined, but it can be measured, according to this interpretation, from the position of the $\frac{1}{2}(^2\Sigma) - \frac{3}{2}(^2\Pi)$ system, whose 0,0 band ought to lie at about 10,000 Å., which is more than 1000 Å. further than the limit reached by Watson and Simon. Of course in true case-c coupling a $^2\Pi$ separation has no meaning, but in partial case-c the levels ($\frac{1}{2}$) and ($\frac{3}{2}$) should have a separation which is related to that which would occur for good Λ - Σ coupling.

§ 6. POTENTIAL-ENERGY CURVES AND DISSOCIATION ENERGY OF PbH

In figure 2 is sketched the probable course of the potential-energy curves of the PbH states involved in this paper. They are all the case-c equivalents of those levels arising from the electron configurations $6p\sigma^26p\pi$, $^2\Pi$ ground state and $6p\sigma6p\pi^2\ ^2\Delta$, $^2\Sigma^+$, $^2\Sigma^-$. Direct extrapolation of the vibrational levels of the

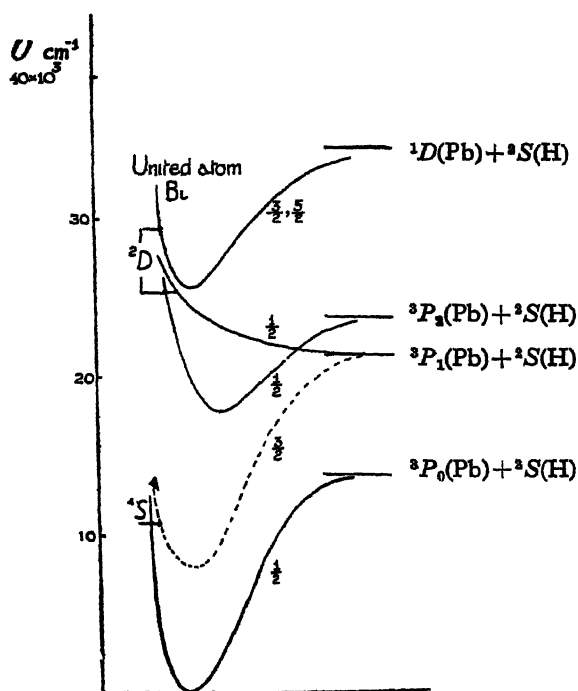


Figure 2. Potential-energy curves for PbH.

ground state gives $D'' = 18,000 \text{ cm.}^{-1}$, which is probably an over-estimate, and so a value of $14,000 \text{ cm.}^{-1}$ has been chosen.

The ground state dissociates into $^3P_0(\text{Pb}) + ^3S(\text{H})$ and the $\frac{3}{2}$ level into $^3P_1 + ^3S$, which fixes D' for the $\frac{3}{2}$ state at about the same value as D'' . The

upper level $\frac{1}{2}$ of the red system must give 3P_2 (Pb) on dissociation to account for its stability and also for the intersection by the curve responsible for its pre-dissociation. According to Watson and Simon, this takes place at a point 3200 cm^{-1} above the $v'=0$ level, i.e. at about 21,200 cm^{-1} , in good agreement with the value of the asymptote of the $\frac{3}{2}$ level, 22,000 cm^{-1} . This repulsive level must be a $\frac{1}{2}$ level, as $^3P_1 + ^3S$ can only give one $\frac{3}{2}$ state, and this has already been accounted for.

The remaining unresolved levels, $\frac{3}{2}$ and $\frac{5}{2}$, must dissociate into 1D (Pb) and have a D value of about 9200 cm^{-1} . All the levels in the figure are probably correlated with the states 4S and 2D of the united atom Bi as shown.

§ 7. THE GeH SPECTRUM

Although no study has yet been made of the GeH spectrum, a comparison with the molecules lying on each side of GeH in figure 1 enables the following predictions to be made about it:—

(a) Its $^2\Pi$ ground state should have a doublet separation of about 920 cm^{-1} —estimated from the 3P ground state of Ge, and also by another method to be described in the next section.

(b) The $^2\Delta$ — $^2\Pi$ system should be the strongest, the $^2\Delta$ level lying midway between the $^2\Delta$ for SiH and SnH at 24,400 cm^{-1} . This places the 0,0 bands at about 4100 Å. and 4260 Å.

(c) The ground state ω'' value should be about 1700 cm^{-1} .

(d) The problem of the $^2\Sigma$ state is somewhat more difficult to settle. No such state has as yet been observed in SiH, but its appearance in SnH suggests that it probably exists in GeH, particularly as Ge is perhaps more closely related to Sn than to Si. Its position is then probably as in figure 1, at about 16,000 cm^{-1} .

§ 8 DOUBLET SEPARATIONS IN HYDRIDES AND HALIDES

It has been established that the ground state of the PbF halide-type group molecules is $^2\Pi$, and the doublet separation has been determined in a number of cases. These are shown in table 3 along with the corresponding hydride

Table 3. $^2\Pi$ separations of CH and CF group type molecules

		H	F	Cl	Br	I
C	^a 28	29	(60)	96	(150)	(250)
Si	149	124	161	208	408	(600)
Ge	940	(920)	(940)	975	1150	(1400)
Sn	2285	2183	2317	2360	2467	(2600)
Pb	7800	(8200)	8266	(8310)	(8450)	(8600)

(predicted value)

value, and also the a value of the atomic p electron, which is certainly involved in the case of the hydride. It will be noticed that the hydride seems to belong to the same family of group molecules as the halides, in which the separation increases as the molecule gets bigger. This is understandable in terms of electron configurations, for the extra eight electrons of the F atom can easily form closed shells in the molecule, leaving the odd one to occupy the same type of orbital as in the hydride.

From the observed separation of 8266 cm.^{-1} for the $^3\Pi$ ground state of PbF , a reasonable value of about 8230 cm.^{-1} seems likely for PbH . This is how the value given in table 1 for PbH was estimated. In the same way, assuming the same general increase in the separation from C to Pb as each halide increases in weight, all the values enclosed in brackets in table 3 have been determined. That for GeH has been compared with the 940 cm.^{-1} obtained from the L-S coupling coefficient of $\text{Ge } ^3P$, and the mean value of 920 cm.^{-1} was used in the discussion on GeH in the previous section. The predicted separations in table 3 are obviously more reliable for the hydrides and fluorides than those for the other halides.

REFERENCES

- HOWELL, H. G., 1943. *Proc. Roy. Soc. A*, **182**, 95
WATSON, W. W., 1938. *Phys. Rev.* **54**, 1068.
WATSON, W. W. and SIMON, R., 1939 *Phys. Rev.* **55**, 358, 1940 *Ibid.* **57**, 708

A PROBLEM OF HEAT CONDUCTION WITH SPHERICAL SYMMETRY

By R. P. BELL, F.R.S.,
Balliol College, Oxford

MS. received 28 September 1944

ABSTRACT. Solutions are given for a problem in heat conduction through two concentric spheres, suggested by the heating up of a body in a furnace. Exact solutions are given, and also simple approximate solutions which will often be applicable in practice

§ 1. INTRODUCTION

THE problem is suggested by the heating up of a body placed at the centre of a furnace cavity. The system considered consists of a conducting sphere of radius a concentric with a spherical cavity of radius b , the annular space being filled by a conducting medium. The outside surface of the cavity is maintained at a fixed temperature, taken arbitrarily as zero, and the initial temperature of the sphere is θ_0 . In practice, the initial temperature

of the annular space will be somewhere between v_0 and zero. We shall give results for these two extreme cases, which will be referred to as assumptions A and B respectively.

The following notation is employed:—

Index 1 refers to $0 < r < a$,

Index 2 refers to $a < r < b$,

K = thermal conductivity, c = specific heat, ρ = density,

$\kappa = K/c\rho$ = diffusivity, $\mu = (\kappa_2/\kappa_1)^{1/2}$, $\sigma = K_2/K_1\mu$, $u = vr$.

The equations to be solved are then

$$\frac{\partial u_1}{\partial t} = \kappa_1 \frac{\partial^2 u_1}{\partial r^2}, \quad \frac{\partial u_2}{\partial t} = \kappa_2 \frac{\partial^2 u_2}{\partial r^2}, \quad \dots \dots (1)$$

subject to the continuity conditions

$$u_1 = u_2, \quad K_1 \left(a \frac{\partial u_1}{\partial r} - u_1 \right) = K_2 \left(a \frac{\partial u_2}{\partial r} - u_2 \right) \quad \text{at } r = a \quad \dots \dots (2)$$

and the initial and boundary conditions

$$\left. \begin{aligned} u_1 &= 0, & r &= 0; & u_2 &= 0, & r &= b; \\ u_1 &= v_0 r, & t &= 0; & u_2 &= v_0 r \text{ (assumption A),} & t &= 0; \\ & & & & \text{or } u_2 &= 0 \text{ (assumption B),} & t &= 0. \end{aligned} \right\} \dots \dots (3)$$

§ 2. GENERAL SOLUTION

The derivation of the solutions will not be given: they were obtained by the usual Fourier method, and in most cases checked by the Laplace transformation method. The exact solutions are:—

Assumption A:

$$v_1 = \frac{2\sigma b v_0}{r} \sum_{n=1}^{\infty} \frac{\sin \mu \alpha_n a \sin \alpha_n (b-a) \sin \mu \alpha_n r e^{-\kappa \alpha_n^2 t}}{\mu \alpha_n a \sin^2 \alpha_n (b-a) + \sigma \alpha_n (b-a) \sin^2 \mu \alpha_n a + \frac{\mu \sigma - 1}{\mu \alpha_n a} \sin^2 \mu \alpha_n a \sin^2 \alpha_n (b-a)};$$

Assumption B:

$$v_1 = \frac{2\sigma v_0}{r} \sum_{n=1}^{\infty} \frac{\sin^2 \alpha_n (b-a) \left\{ \frac{1}{\mu \alpha_n a} \sin \mu \alpha_n a - \cos \mu \alpha_n a \right\} \sin \mu \alpha_n r e^{-\kappa \alpha_n^2 t}}{\mu \alpha_n \sin^2 \alpha_n (b-a) + \sigma \alpha_n (b-a) \sin^2 \mu \alpha_n a + \frac{\mu \sigma - 1}{\mu \alpha_n a} \sin^2 \mu \alpha_n a \sin^2 \alpha_n (b-a)}, \quad \dots \dots (4)$$

where α_n is the n th positive root of the equation

$$\sin \alpha (b-a) \{ \mu \alpha a \cos \mu \alpha a - \sin \mu \alpha a \} + \mu \sigma \sin \mu \alpha a \{ \alpha a \cos \alpha (b-a) + \sin \alpha (b-a) \} = 0 \quad \dots \dots (5)$$

The solution above is complete only if $\mu a/(b-a)$ is irrational. If $\mu a/(b-a) = p/q$, where p and q are integers, then equation (5) admits roots such that $\alpha \mu a$ and $\alpha (b-a)$ are integral multiples of π , and the corresponding terms (which are limiting forms of those already given) take the forms;

Assumption A:

$$-\frac{2b\sigma v_0}{\pi r(p\sigma + q)} \sum_{s=1}^{\infty} \frac{(-)^{(p+q)s}}{s} \sin \frac{ps\pi r}{a} \exp. \left\{ -\frac{s^2 p^2 \pi^2}{\mu^2 a^2} \kappa_2 t \right\};$$

Assumption B:

$$-\frac{2a\sigma v_0}{\pi r(p\sigma + q)} \sum_{s=1}^{\infty} \frac{(-)^{sp}}{s} \sin \frac{ps\pi r}{a} \exp. \left\{ -\frac{s^2 p^2 \pi^2}{\mu^2 a^2} \kappa_2 t \right\}.$$

.....(6)

The problem corresponding to assumption A has previously been treated by Carslaw (1920) in connection with the rate of cooling of the earth's crust. Allowing for differences in notation, his formula for the additional terms agrees with our equation (6A), but his version of (4A) has an incorrect denominator, which is not of the right physical dimensions.

§ 3. PARTICULAR CASES

For practical purposes these equations can often be considerably simplified. Thus, if the central sphere is a sufficiently good conductor, the temperature inside $r=a$ can be taken as uniform, i.e. v_1 is independent of r . Under these conditions the second continuity condition in equation (2) is replaced by

$$3K_2 \frac{\partial v_2}{\partial r} = ac_1 p_1 \frac{\partial v_1}{\partial t} \text{ at } r=a. \quad \text{.....(7)}$$

This leads to the following equations:

Assumption A:

$$v_1 = \frac{6v_0 b}{a} \sum_{n=1}^{\infty} \frac{\sin \alpha_n (b-a) e^{-\kappa_n \alpha_n^2 t}}{2 \frac{\mu}{\sigma} \alpha_n a \sin^2 \alpha_n (b-a) + 3 \alpha_n (b-a) - \frac{3}{2} \sin 2 \alpha_n (b-a)};$$

Assumption B

$$v_1 = 2v_0 \sum_{n=1}^{\infty} \frac{\alpha_n a \sin^2 \alpha_n (b-a) e^{-\kappa_n \alpha_n^2 t}}{2 \alpha_n a \sin^2 \alpha_n (b-a) + 3 \frac{\sigma}{\mu} \{ \alpha_n (b-a) - \frac{1}{2} \sin 2 \alpha_n (b-a) \}},$$

.....(8)

where α_n is now the n th positive root of the equation

$$\alpha_n a \cot \alpha_n (b-a) + 1 = \frac{1}{3} \frac{\mu}{\sigma} \alpha_n^2 a^2. \quad \text{.....(9)}$$

In this case there are no terms analogous to (6). It may be noted that when $a=0$, (8B) reduces to

$$v_1 = 2v_0 \sum_{n=1}^{\infty} \left(\frac{-}{-} \right)^n e^{-\frac{n^2 \pi^2}{b^2} \kappa_2 t}, \quad \text{.....(10)}$$

agreeing with the well-known result for the temperature at the centre of a sphere, the surface of which is held at a constant temperature.

In the practical problem of a body heating up in a furnace, it is often the later stages of the process which are of interest; for example, we may wish to

know what time is necessary for the whole of the body to reach the furnace temperature within certain limits of error. Under these circumstances it is a good approximation to take only the first term of equation (4) or (8), and the additional terms in (6) are negligible. Moreover, for the later stages it is easy to find the conditions under which it is legitimate to assume a uniform temperature within the sphere $r=a$. If v_c and v_a are respectively the temperatures at the centre and surface of the sphere in the later stages, then equations (4) and (5) give (for both assumptions A and B)

$$1 - \frac{v_a}{v_c} = 1 - \frac{\sin \mu \alpha_1 a}{\mu \alpha_1 a} \leq \frac{1}{2} \sigma \mu \frac{b}{b-a}. \quad \dots\dots(11)$$

The condition for using the first term of (8) rather than the first term of (4) is thus $\frac{1}{2} \sigma \mu b / (b-a) \ll 1$. Finally, if the following conditions hold simultaneously,

$$\frac{1}{2} \sigma \mu \frac{b}{b-a} \ll 1, \quad \frac{3\sigma}{\mu} \cdot \frac{b(b-a)}{a^2} \ll 1, \quad \dots\dots(12)$$

then either assumption A or assumption B gives the following very simple formula for the later stages:

$$v_1 = v_0 \exp. \left\{ - \frac{3\sigma}{\mu} \cdot \frac{b}{a^2(b-a)} \kappa_2 t \right\}. \quad \dots\dots(13)$$

Equation (13) will often be applicable in practice, especially if a high accuracy is not needed. The proportional errors in the times which it predicts for a given value of v_1/v_0 will be of the same order of magnitude as the two quantities in equation (12). Equation (13) will also give approximate values for non-spherical systems, provided that average values of $b/a^2(b-a)$ are taken.

If the furnace temperature is high and the annular space wide, heating by radiation may be more important than heating by conduction. If the temperature inside the sphere is effectively uniform, then heating by radiation alone would give

$$v_1 = v_0 e^{-3Et/(\omega_1 v_1)}, \quad \dots\dots(14)$$

where the emissivity E can be estimated from the Stefan radiation constant and the absolute temperature of the formula. The result for non-uniform temperature inside the sphere has been given by Awbery (1927) in connection with the cooling of apples in refrigerators

REFERENCES.

- AWBERY, J. H., 1927. *Phil. Mag* 4, 629.
CARSLAW, H S, 1920 *Proc Camb Phil. Soc.* 20, 399.

COLORATION AND LUMINESCENCE PRODUCED BY RADIUM RAYS IN THE DIFFERENT VARIETIES OF QUARTZ, AND SOME OPTIC PROPERTIES OF THESE VARIETIES

BY S. P. CHOONG

(in orthodox romanization, Chung Shêng-Piao),
Institute of Radium, National Academy of Peiping, Kunming, China

Communicated by W Band ; MS. received 25 August 1944

ABSTRACT. Systematic studies of the coloration and luminescence produced by radium rays, on the one hand, and of the absorption spectra and ultra-violet transparency on the other, have been made on four varieties of quartz, namely, the colourless, the smoky, the citrine and the rose, in their crystalline and vitreous states. The chief results may be summarized as follows:—(1) After being exposed to the irradiation of the β and γ rays of radium, the crystals of all four varieties of quartz turn to different shades of black, so that in appearance they resemble very smoky quartz. The persistence of the coloration under heat treatment depends, however, on the variety. (2) The same rays render vitreous quartz of the four varieties violet in colour instead of smoky. (3) Under the influence of radium rays, all the specimens of quartz, crystalline and vitreous, fluoresce and thermophosphoresce in bluish-green or bluish-white light, with the sole exception of crystalline rose quartz, which emits orange light. (4) Under heat treatment, coloration and thermophosphorescence do not disappear simultaneously, the latter being more persistent. (5) Complementary observations with the use of X rays and ultra-violet radiations have shown that the coloration and the luminescence caused by X rays are exactly the same as those produced by the radium rays, and that the ionizing ultra-violet radiations have a very marked decolorizing effect on the smoky quartz, but have no similar effect on the artificial violet colour of vitreous quartz. (6) Among the naturally and artificially coloured specimens of quartz, the violet vitreous quartz is the only one which possesses a band structure in its ultra-violet absorption spectrum. (7) In passing from the naturally colourless quartz to the three decolorized varieties, the limit of ultra-violet transparency recedes from $\lambda=1850$ A. to the vicinity of $\lambda=2250$. Among the decolorized specimens of the three coloured varieties, crystalline and vitreous quartz of the smoky variety are the most transparent and the rose variety the least.

Certain discrepancies, concerning the coloration and luminescence of quartz, between the present results and those of previous workers are pointed out and discussed.

§ 1. INTRODUCTION

THE coloration and luminescence of quartz produced by the rays of radium have already been the subject of study by several investigators, notably A. Bensaude and G. Costanzo (1922), S. C. Lind and D. C. Bardwell (1923), and J. Hoffmann (1931). Their work was mainly limited to the observations of the effects produced either on colourless quartz (crystalline and vitreous) or on certain coloured crystalline quartzes. Since the phenomena of coloration and luminescence produced in quartz are, as in other transparent minerals,

complicated in nature, it seems necessary to follow closely the variations of these phenomena with the physical states of the quartz, and also of the impurities it contains, in order that a better understanding may be acquired. In addition, a knowledge of certain thermal and optical properties of the coloured and luminescent substances appears also to be essential. It was such considerations as these that led the writer to study systematically (1) the effects of coloration and luminescence caused by the radium rays in different varieties of quartz, both crystalline and vitreous, and (2) the spectroscopic properties of quartz before and after heating, using both naturally coloured specimens and others artificially coloured by radium rays. With a view to comparing such effects with those caused by other kinds of rays, some complementary observations, with the use of x rays and ultra-violet radiations as exciting sources, have also been carried out.

§ 2. EXPERIMENTAL

Four varieties of quartz, namely, the colourless, the smoky, the citrine, and the rose, were used in the investigation. The specimens of vitreous quartz of each variety were prepared by fusing the corresponding crystals in a gas-oxygen flame.

50 mg. of radium contained in a platinum tube, of which the wall was 0.5 mm. in thickness, were used in the experiments. Owing to the absorption of the tube wall, it is certain that the colouring and luminescent effects observed are due entirely to the external-conversion β rays and the penetrating γ rays. The α rays, and almost all of the β rays, emitted directly from the radium and its derivatives are stopped by the wall of the tube. The specimens of quartz to be irradiated were placed on the radium tube. The time of irradiation varied from 30 to 50 days.

As the source of x rays, a Roentgen tube with copper anode and aluminium-foil window was used; at a voltage of 50 kv. and a current of 10 ma., the time of exposure was 4 hours.

The ultra-violet radiation used for the study of the colouring and luminescent effects, and also for the study of absorption spectra of the specimens of coloured quartz, was furnished by a hydrogen discharge tube. This tube, which had a power of about 1 kw., was closed at one end by a quartz window and at the other end by a fluorite window. The specimen to be irradiated was put in contact with one of these windows, in order to minimize the loss of radiations of short wavelength due to air absorption. The extent of the spectral region used could be varied according to the kind of window with which a specimen was placed in contact.

A large quartz spectrograph of the Littrow model was used for the study of the absorption spectra and a fluorite vacuum spectrograph for the comparison of the transparency limits of the colourless and decolourized specimens. The spark spectrum of silver was used for this comparison.

The photographic plates used for the ultra-violet region of short wave-lengths were ordinary plates sensitized with sodium salicylate, according to the process described by J. Terrien (1936).

§ 3. RESULTS

(a) *Coloration*

When a specimen of quartz a few mm. in thickness is exposed to the rays of radium, the part of its surface in contact with or near the radium tube becomes rapidly coloured. Evidently this is due to β rays, which are known to be much more effective in colouring and much less penetrating than the γ rays. After a relatively short period, the coloration of the superficial layer attains saturation, while that of the inner layers continues to intensify until, finally, the whole mass becomes sensibly uniform in colour.

(1) *Crystalline quartz.* The radium rays deepen the smoky appearance of the smoky quartz and transform the citrine and the rose quartz to smoky. The same effect is also observed in specimens of which the natural colours have previously been eliminated by heat treatment, while in the case of natural colourless quartz the coloration is generally not strong and lacks homogeneity.

(2) *Vitreous quartz.* Vitreous quartz specimens prepared from crystals of the different varieties are transparent and colourless. The radium rays give a violet colour to these vitreous bodies, instead of smoky. The maximum possible degree of coloration differs with the variety of quartz. In general, the effect is much more prominent in coloured quartz than in the colourless. The coloration in the latter is often irregular; in most cases the coloured and the unaffected parts are separated by a sharp boundary.

(b) *Decoloration*

The natural colours and those produced by the radium rays in quartz, both crystalline and vitreous, can be eliminated more or less easily by heating. The artificial smoky and violet tints fade with a perceptible rapidity at 190° c. and 300° c. respectively, and very rapidly at 400° c. Natural smoky coloration is more persistent than artificial, and begins to fade at about 220° c. The citrine quartz does not lose its yellow colour until 500° c. The rose-coloured quartz is the most persistent of all; complete decoloration cannot be attained below 700° c.

After the colorations produced by the radium rays have been eliminated by heating, the specimens regain their original transparency and they can be coloured again. Apparently the effect is reversible and without fatigue.

It is interesting to point out that the naturally coloured crystals of quartz, after having been blackened by the radium rays, may have their original colours completely restored by a suitable heat treatment. This fact indicates that the radium rays do not actually destroy natural colours, but that these are simply masked by superimposed smokiness.

(c) *Thermophosphorescence*

The varieties of quartz under investigation do not originally possess the property of emitting light on heating. However, under the action of radium rays they readily acquire the property of exhibiting thermophosphorescence. Crystalline rose quartz emits orange light at high temperatures, and its vitreous state gives bluish-green light. Both the crystalline and vitreous specimens of the other varieties emit either bluish-green or bluish-white light.

The specimens of quartz which have been well excited by radium rays phosphoresce brilliantly at high temperatures. It may be emphasized that thermophosphorescence and the artificial colours do not disappear simultaneously during heating. For example, at 400°C . the former lasts generally some ten minutes, while the latter disappears often in the first minutes of heating. Because of this difference in persistence, it is possible to eliminate coloration without completely destroying thermophosphorescence. The time required for the decoloration depends greatly on the temperature and also on the nature of the colour. If a strongly coloured specimen of vitreous quartz is heated at 270°C . instead of at 400°C ., it is necessary to prolong the heating from the order of one minute to about two hundred hours for the removal of the violet coloration. In spite of such prolonged heating, it was found that after the heat treatment there remains always a notable proportion of thermophosphorescence. Some observations performed at 200° show that the violet fused quartz loses neither its colour nor its emissive power of phosphorescence even after a heating of ten days, and that, on the other hand, the artificial smoky specimen loses its coloration in several hours, although its thermophosphorescence does not appear to be modified sensibly. The above facts seem to be important, since it is generally believed that the coloration and the thermophosphorescence of a substance disappear simultaneously during heating. This question will be brought up again and discussed later in this paper.

(d) *Fluorescence*

Specimens of all four varieties of quartz fluoresce more or less strongly when they are exposed to the β and γ rays of radium. The fluorescence is either bluish green or bluish white in colour, except that from the crystalline rose quartz, which emits an orange light, as in the case of thermophosphorescence. In general, the fluorescence of quartz is stronger in intensity for the vitreous than for the crystalline state.

(e) *Effects produced by x rays and ultra-violet radiations*

Studies of the colouring and luminescence produced in quartz by radium rays have led the writer to examine briefly the cases when x rays and ultra-violet radiations are employed.

It was found that coloration and luminescence caused by x rays are in every respect similar to those caused by β and γ rays.

The effect of coloration produced by ultra-violet radiation is different from that produced by radium rays and x rays. Thus crystals of smoky quartz are rendered brownish but not smoky. With the specific type of hydrogen tube used, a period of exposure of the order of thirty hours is necessary for this colour to be perceptible, while no coloration was observed, for the same duration of irradiation, in the case of vitreous smoky quartz and the crystalline and vitreous specimens of other varieties.

In the course of the experiments on coloration of quartz by ultra-violet radiation, an interesting phenomenon was observed, viz., that these radiations are capable of removing the smoky colour caused by radium rays. The effect appears to be fairly strong, since an irradiation of only thirty to forty minutes

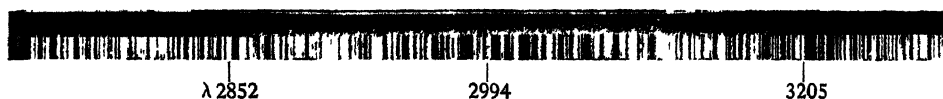


Figure 1 Absorption band of violet fused quartz.

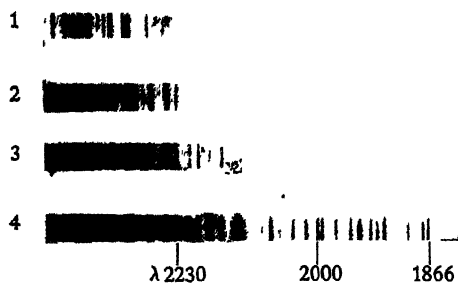


Figure 2 Spark spectra of silver, showing the limits of transparency for the crystals of the varieties: (1) rose, (2) yellow, (3) smoky, (4) colourless

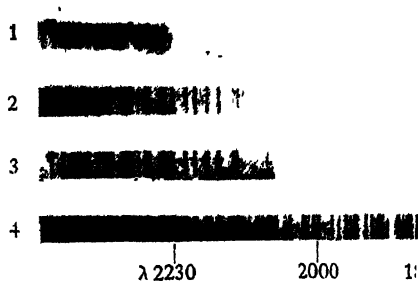


Figure 3 Spectra of silver, showing the limits of transparency for the vitreous crystals of the varieties (1) rose, (2) yellow, (3) smoky, (4) colourless

is sufficient to eliminate the colour. On the violet colour produced by radium rays in vitreous quartz, ultra-violet radiations show no perceptible effect.

The thermophosphorescence caused by the ultra-violet radiations is essentially the same as that due to radium rays. Apparently fused quartz acquires equally well the property of thermophosphorescence whether it is placed under the radiations which come through the quartz window of the hydrogen tube or from the fluorite window, whereas in the case of crystalline quartz this is no longer true; thermophosphorescence can be produced only by the radiations which pass through the fluorite window. This fact indicates that for the crystalline quartz the effect of thermophosphorescence is due to the ultra-violet radiations of wave-lengths inferior to 1850 Å.

(f) *Spectroscopic properties*

(1) *Absorption spectra of coloured quartz.* The absorption spectra of coloured quartz have been studied in the visible and ultra-violet regions. For specimens of naturally coloured quartz and those rendered smoky by radium rays, such spectra show no band structure in the regions of interest. The fused quartz coloured violet by radium or x rays was found to be the only specimen with an absorption band. This band, which, so far as the writer knows, has not been observed before, is situated in the interval between $\lambda = 2750$ and $\lambda = 3200$, with maximum absorption at about $\lambda = 3000$ Å., as shown in the plate, figure 1.

(2) *Transparency limits of the decolourized quartz.* In connection with the absorption spectra of coloured quartz it may be interesting to compare the ultra-violet transparency (or rather the limit of transparency) of the decolourized specimens of coloured varieties with that of naturally colourless quartz. Such a comparison has been made for specimens in crystalline as well as in vitreous states. Corresponding crystalline and vitreous specimens were taken from the same pieces of quartz in order to avoid difference in the quantities of impurities. The specimens employed were all 1 cm. in thickness. Figures 2 and 3 represent the spectrograms for the specimens in crystalline and vitreous states respectively.

It can be seen from the above figures that the decolourized specimens of the coloured varieties are much less transparent than colourless quartz in the ultra-violet region of short wave-lengths. In passing from the naturally colourless quartz to the decolourized quartz of the different varieties, the limits of transparency receded from $\lambda = 1850$ to the vicinity of $\lambda = 2250$. Among the decolourized specimens, smoky quartz is the most transparent and rose the least. Evidently the limit of transparency varies with the quantity of impurities; consequently it will differ from sample to sample for a given variety. The values indicated in the above figures, representing the limit of transparency of some particular samples, simply give orders of magnitude.

§ 4. DISCUSSION OF RESULTS

The preceding investigations permit us to recognize some regularities and some anomalies concerning the coloration and luminescence produced by certain kinds of radiations in the different varieties of quartz. Equally they permit us to draw some conclusions as to the agencies which give rise to the coloration of quartz. With these experimental facts it is possible to clear up certain

problems regarding the nature of coloration and luminescence produced by ionizing rays and to the connexion between the phenomena of coloration and of thermophosphorescence.

E. F. Holden (1925) has compared the colour of the smoky quartz of natural origin with that produced in the laboratory by radium rays. According to him, there is a possibility that the coloration of the smoky quartz is due to the atoms of silicon, liberated under the action of the rays of radioactive substances during long geological epochs. This consideration appears to be improbable for the reason that the smoky colour produced in the laboratory is incontestably less persistent, both under heat treatment and under exposure to ultra-violet radiations, than natural smoky coloration. The facts that the crystalline and the vitreous colourless quartz are generally much lighter coloured than other varieties, and that there are often some specimens which do not allow themselves to be influenced at all by the ionizing rays, appear to indicate that neither the artificial smoky colour nor the artificial violet colour of the vitreous quartz can be attributed to the atoms or ions of silicon. J. Hoffmann (1931) referred the violet colour of the vitreous quartz to ferrous and ferric ions in the presence of titanium or zirconium. According to this author it is unlikely that the colour of the fused quartz is due to the ions or atoms of manganese, although the soda-calcic glasses, which contain some manganese, also become violet when irradiated by radium rays. It may be mentioned that the characteristic absorption band of the violet fused quartz is unfortunately found in the ultra-violet absorption region of the glasses in question; otherwise it would be possible to decide by spectroscopic method, without ambiguity, whether the absorbing agents in the glasses and in the fused quartz are different in nature.

Lind and Bardwell (1923) have described briefly their observation on the fluorescence of quartz. According to them, quartz does not fluoresce under radium rays. However, the writer has not been able to confirm this result; for, without exception, all the specimens of quartz examined fluoresce more or less strongly when they are irradiated by the β and γ rays of radium.

In comparing the spectrograms for any one variety of quartz in figures 2 and 3, it is seen that the crystals are generally somewhat more transparent than the vitreous material, and that this difference in transparency is exceptionally pronounced in the case of yellow quartz. A slight variation in transparency arising from the change in the physical state appears to be imputable to the effect of molecular scattering, which, as we know, is somewhat stronger in the vitreous than in the crystalline state. However, the unusually large variation of transparency of the yellow quartz seems to indicate that the nature of the absorbing impurities has been fundamentally modified by the change of state.

It has long been known that the ionizing rays can produce the effects of coloration and thermophosphorescence in many minerals, salts and glasses. Since these two phenomena are frequently coexistent and disappear simultaneously during heating, it would be quite natural to suppose a common basis. The theory of internal photoelectric effect has been generally adopted. According to this theory, certain electrons of the molecules which absorb the energy of the ionizing rays are displaced to some metastable levels, where they produce (or sometimes may not produce) a coloration. Their return to the

normal levels, following, for instance, a thermal agitation, involves an emission of the phosphorescence and loss of coloration. Obviously this theory associates the phenomenon of coloration with that of thermophosphorescence. However, it has been pointed out by S. C. Lind (1920) that this theory is not always true. He gave as examples the cases of certain glasses and of fused quartz. According to him, the violet colour produced in these two substances by the radium rays does not disappear below 500°C ., and on the other hand, their thermophosphorescence can be eliminated at 200°C . The observations of Lind on fused quartz are evidently in contradiction to the results of the present investigation, since it is the thermophosphorescence of quartz, and not the coloration, which is more persistent.

This inconsistency in observation induced the writer to make a careful comparison of the thermophosphorescence and the decolourization due to heating violet fused quartz and violet glass, both having been coloured by radium rays. The results are summed up as follows.—(1) Under the same experimental conditions, thermophosphorescence of vitreous quartz is incontestably stronger in intensity than that of glass. (2) In conformity with the result of Lind, thermophosphorescence of glass disappears rapidly at 200°C ., but its violet colour remains practically unchanged. In the case of fused quartz, as mentioned above, neither its violet colour nor its thermophosphorescence is sensibly modified at this temperature; (3) violet glass begins to decolourize at 350°C ., a temperature some 150° lower than that found by Lind. On heating the violet fused quartz at 270°C . for some two hundred hours, its colour is completely eliminated, but it still shows notable thermophosphorescence. These observations certainly disclose an essential difference between the properties of violet fused quartz and those of violet glass in respect of thermophosphorescence and of thermal decoloration.

ACKNOWLEDGMENTS

The writer wishes to express his indebtedness to Dr. S. S. Lu, whose x-ray set was freely used in the experiments, and to Mr. Chang Yuan-Lung for his assistance in experiments with quartz

REFERENCES

- BENSAUDE, A and COSTANZO, G, 1922 *J Phys (Chim)*, **3**, 384.
HOFFMANN, J., 1931 *Z. anorg Chem* **197**, 29
HOLDEN, E F, 1925. *Am. Mineral* **10**, 203
LIND, S. C., 1920 *J. Phys. Chem* **24**, 442.
LIND, S. C and BARDWELL, D C, 1923 *J Franklin Inst* **196**, 381.
TERRIEN, J., 1936 *C.R. Acad Sci., Paris*, **202**, 211.

THE POLARIZING ANGLE FOR REFLECTION AT THE BOUNDARY BETWEEN TWO ABSORBING MEDIA

BY L. PINCHERLE,
King's College, London

MS. received 28 August 1944

ABSTRACT. The condition for the absence of a reflected wave when a plane (inhomogeneous) electromagnetic wave is incident upon the plane boundary between two absorbing media is

$$\cot \phi_1 \cot \psi_2 + \cot \psi_1 \cot \phi_2 = 2,$$

ϕ_s, ψ_s being the angles that the planes of equal phase and the planes of equal amplitude respectively make with the boundary.

For small values of the conductivities no reflection is obtained at the following angles :

$$\begin{aligned} \tan \phi_1 &= \sqrt{E}, & \tan \phi_2 &= \frac{1}{\sqrt{E}}, \\ \tan \psi_1 &= \frac{1 + E \frac{\eta_1}{\eta_2}}{\sqrt{E} \left[\left(2 + \frac{1}{E} \right) \frac{\eta_1}{\eta_2} - 1 \right]}, \\ \tan \psi_2 &= \frac{1 + E \frac{\eta_1}{\eta_2}}{\sqrt{E} \left(E + 2 - \frac{\eta_1}{\eta_2} \right)}, \end{aligned}$$

where

$$E = \frac{\epsilon_2}{\epsilon_1}; \quad \eta_1 = \frac{\sigma_1}{\omega \epsilon_1}; \quad \eta_2 = \frac{\sigma_2}{\omega \epsilon_2}$$

and ϵ is the dielectric constant, σ the conductivity, and ω is $2\pi \times$ frequency

§1. INTRODUCTION

THE problem of determining theoretically under what conditions there is no reflected wave when a plane electro-magnetic wave is incident upon the plane boundary between two different media is a well-known one.* However, the analogue of Brewster's formula in the case of absorbing media has not, to my knowledge, been given explicitly. It is the purpose of this paper to establish it.

It is possible that the theoretical results obtained may have practical applications: dielectric constants and conductivities could be determined by measuring the angles at which no reflection occurs, in the same way as refractive indices of perfect dielectrics are found by measuring the Brewster angle.

* See, for example, J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Co., 1941), p. 516; J. C. Slater, *Microwave Transmission* (McGraw-Hill Book Co., 1942), p. 117, or W. König, *Handbuch der Physik*, Bd. xx (Springer, 1928), p. 194.

§ 2. FOUNDATION OF THE EQUATIONS

We take the boundary of the two media as the plane yz . We assume the electric vector of the waves to be parallel to xz , the plane of incidence, as this is the case in which we expect the reflected wave to be absent; we indicate by ϵ_1, σ_1 the dielectric constant and the conductivity of the first medium, and by ϵ_2, σ_2 those of the second; we assume the magnetic permeability of both media to be that of empty space. We indicate by A_s, B_s ($s=1, 2$) the real and imaginary parts of the propagation constants in the two media; by ϕ_1 the angle of incidence, namely, the angle that the normal to the planes of equal phase in the first medium makes with the positive x -axis; by ϕ_2 the similarly defined angle of refraction; and by ψ_1, ψ_2 the angles that the normals to the plane of equal amplitude make with the x -axis in the two media.

In absorbing media, the planes of equal phase and the planes of equal amplitude do not in general coincide, i.e., the waves are inhomogeneous, and the introduction of ψ_1, ψ_2 , as well as ϕ_1, ϕ_2 , is necessary. Our wave will thus be represented by

$$e^{-A_s \cos \psi_s x - A_s \sin \psi_s z - i(B_s \cos \phi_s x + B_s \sin \phi_s z) + i\omega t} \quad (s=1, 2), \quad \dots\dots (1)$$

where $\omega = 2\pi \times \text{frequency}$.

Such waves can be generated in several ways: the simplest is by refraction from air to an absorbing dielectric. Waves in hollow pipes filled with an absorbing medium, and all kinds of cylindrical and spherical waves, can be represented by the superposition of inhomogeneous plane waves.

Even in perfect dielectrics the angles ϕ, ψ may be different, as, for instance, in the field in the second medium, when total reflection takes place. In this case, the planes of equal amplitude are at right angles to those of equal phase.

In this discussion, rationalized M.K.S. units are used.

The quantities in equation (1) are connected by various relations. Firstly, by the formulae defining A_s, B_s , viz.

$$\left. \begin{aligned} B_s^2 - A_s^2 &= \epsilon_s \mu \omega^2, \\ \text{and} \quad 2A_s B_s \cos(\psi_s - \phi_s) &= \mu \sigma_s \omega; \end{aligned} \right\} \quad \dots\dots (2)$$

secondly, by the law of refraction,

$$\left. \begin{aligned} B_1 \sin \phi_1 &= B_2 \sin \phi_2, \\ \text{and} \quad A_1 \sin \psi_1 &= A_2 \sin \psi_2; \end{aligned} \right\} \quad \dots\dots (3)$$

and finally by the conditions for the absence of the reflected wave; these are given by Stratton (l.c. p. 517, eq. (3)), and, using our notation, can be written

$$\left. \begin{aligned} B_1 B_2 \cos(\phi_1 + \phi_2) &= A_1 A_2 \cos(\psi_1 + \psi_2), \\ A_1 B_2 \cos(\psi_1 + \phi_2) &= -A_2 B_1 \cos(\phi_1 + \psi_2). \end{aligned} \right\} \quad \dots\dots (4)$$

If we multiply A_s^2, B_s^2, σ_s and ω , by the same constant factor, the angles remain unchanged; it is then convenient to introduce the following dimensionless ratios:

$$E = \frac{\epsilon_2}{\epsilon_1}; \quad \eta_1 = \frac{\sigma_1}{\omega \epsilon_1}; \quad \eta_2 = \frac{\sigma_2}{\omega \epsilon_2}. \quad \dots\dots (5)$$

$$\alpha_1 = \frac{A_1}{R_1}; \quad \alpha_2 = \frac{A_2}{R_2}; \quad \beta = \frac{B_2}{R_2}. \quad \dots\dots (6)$$

Equations (2), (3) and (4) then become :

$$\begin{aligned} \eta_s &= \frac{2\alpha_s \cos(\psi_s - \phi_s)}{1 - \alpha_s^2}; & E &= \beta^2 \frac{1 - \alpha_2^2}{1 - \alpha_1^2}; \\ \beta &= \frac{\sin \phi_1}{\sin \phi_2}; & \frac{\alpha_2}{\alpha_1} &= \frac{\sin \phi_2 \sin \psi_1}{\sin \phi_1 \sin \psi_2}; & \dots\dots(7) \\ \cos(\phi_1 + \phi_2) &= \alpha_1 \alpha_2 \cos(\psi_1 + \psi_2); \\ \alpha_1 \cos(\psi_1 + \phi_2) &= -\alpha_2 \cos(\phi_1 + \psi_2). \end{aligned}$$

η_1 , η_2 , E and β are necessarily positive quantities. If, in addition, we want the amplitude of the waves to decrease in the direction of propagation, which is the physically important case,* we must have

$$\alpha_s \cos(\psi_s - \phi_s) > 0.$$

Also, of course,

$$|\alpha_1|, |\alpha_2| < 1.$$

§ 3. CONDITIONS FOR ABSENCE OF REFLECTION

We have seven equations connecting ten quantities. Usually the constants of the two media, namely E , η_1 , η_2 , will be given, and we shall want to find at what angles there is no reflected wave. Conversely we can ask for what media we find no reflection at given angles. It appears immediately from the last three of equations (7) that the four angles cannot all be given arbitrarily. eliminating α_1 , α_2 , we find that the following relation must be satisfied :

$$\cot \phi_1 \cot \psi_2 + \cot \psi_1 \cot \phi_2 = 2. \quad \dots\dots(8)$$

This, for absorbing media, is the analogue of Brewster's condition. (For $\eta_1 = \eta_2 = 0$ we have $\psi_1 = \phi_1$, $\psi_2 = \phi_2$, and (8) becomes $\cot \phi_1 \cot \phi_2 = 1$, the usual condition.)

Even subject to (8), the physically possible values of the angles are further restricted. Each of them can take any value between $-\pi/2$ and $+\pi/2$, but it is easily seen from (7), (8) that ϕ_1 and ϕ_2 must have the same sign, and we shall take them as positive; then ψ_1 and ψ_2 cannot be both negative, because this would contradict (8). We must therefore have either all four angles positive, or three positive and one of the ψ_s negative. In this last case, the corresponding α_s is negative, so that $\cos(\psi_s - \phi_s)$ must be negative, to satisfy the condition $\alpha_s \cos(\psi_s - \phi_s) > 0$. The sign of α_s is immaterial when $\cos(\psi_s - \phi_s) = 0$; then, unless $|\alpha_s| = 1$, the corresponding medium must be a perfect dielectric, because of the first of equations (7); this case has been treated by Stratton (see later)

Solving the last three equations (7), we find :

$$\begin{aligned} \alpha_1^2 &= \frac{\sin \phi_1 \sin \psi_2 \cos(\phi_1 + \phi_2)}{\sin \psi_1 \sin \phi_2 \cos(\psi_1 + \psi_2)}; \\ \alpha_2^2 &= \frac{\sin \psi_1 \sin \phi_2 \cos(\phi_1 + \phi_2)}{\sin \phi_1 \sin \psi_2 \cos(\psi_1 + \psi_2)}, \end{aligned} \quad \dots\dots(9)$$

and the angles must thus be restricted to such values as make these two quantities positive and less than unity.

* Stratton, l.c., chooses instead to make the amplitude of the incident wave vanish at infinity.

The case $\alpha_1 = \alpha_2 = 1$ corresponds to $\eta_1 = \eta_2$, in which case $\psi_1 = \phi_1$ and $\psi_2 = \phi_2$ (see later).

When the above conditions are satisfied, the equations (7) can be easily solved, and the values of E , η_1 , η_2 found.

§ 4. PARTICULAR CASES

Reverting to the usual case, the equations (7) can be solved for given E , η_1 , η_2 , but, as with the equivalent equations given by Stratton, no simple expressions can be obtained for the angles.

The solution, however, is easily obtained in limiting cases, to which we shall confine ourselves.

(a) $\eta_1, \eta_2 \ll 1$

Keeping only quadratic terms in η_1, η_2 , we have

$$\begin{aligned}\tan \phi_1 &= \sqrt{E} \left[1 + (\eta_2 - \eta_1) \frac{(1-E)\eta_1 + (3E+1)\eta_2}{8(E+1)} \right], \\ \tan \phi_2 &= \frac{1}{\sqrt{E}} \left[1 + (\eta_2 - \eta_1) \frac{(E+3)\eta_1 + (E-1)\eta_2}{8(E+1)} \right], \\ \tan \psi_1 &= \sqrt{E} \frac{\eta_2 + E\eta_1}{(2E+1)\eta_1 - E\eta_2}, \\ \tan \psi_2 &= \frac{1}{\sqrt{E}} \frac{\eta_2 + E\eta_1}{(E+2)\eta_2 - \eta_1}.\end{aligned}\quad (10)$$

ψ_1, ψ_2 depend thus only on the ratio $\eta_1/\eta_2 = \epsilon_2\sigma_1/\epsilon_1\sigma_2$. The values of $\tan \psi_1$ and $\tan \psi_2$ for different values of this ratio are given in the following table:—

η_1/η_2	$\tan \psi_1$	$\tan \psi_2$
0	$-1/\sqrt{E}$	$1/\sqrt{E(E-2)}$
$E/(2E+1)$	$\mp \infty$	$1/2\sqrt{E}$
1	\sqrt{E}	$1/\sqrt{E}$
$E-2$	$\sqrt{E/2}$	$\pm \infty$
∞	$E\sqrt{E/(2E+1)}$	$-\sqrt{E}$

When η_1, η_2 increase, so that the above approximation is no longer valid, ψ_1, ψ_2 depend on the actual values of η_1, η_2 and not only on their ratio. The angles, however, vary much, as they do for η_1, η_2 small.

(b) $\eta_1, \eta_2 \gg 1$

We have

$$\begin{aligned}\tan \phi_1 &= \sqrt{\frac{E\eta_2}{\eta_1}} \left[1 - \frac{\eta_2 - \eta_1}{2\eta_1\eta_2} \right], \\ \tan \phi_2 &= \sqrt{\frac{\eta_1}{E\eta_2}} \left[1 + \frac{\eta_2 - \eta_1}{2\eta_1\eta_2} \right], \\ \tan \psi_1 &= \sqrt{\frac{E\eta_2}{\eta_1}} \left[1 + \frac{\eta_2 - \eta_1}{2\eta_1\eta_2} \right], \\ \tan \psi_2 &= \sqrt{\frac{\eta_1}{E\eta_2}} \left[1 - \frac{\eta_2 - \eta_1}{2\eta_1\eta_2} \right].\end{aligned}\quad (11)$$

Thus in this case all four angles are positive, and ψ_s and ϕ_s differ only by terms in η^{-1} . In this approximation also the angles depend only on the ratio η_1/η_2 .

(c) *Medium (1) is a dielectric and medium (2) is a conductor*

This is the case treated by Stratton (l.c. p. 520). Of course, if medium (1) is a perfect dielectric, the incident wave must have the planes of equal phase perpendicular to the planes of equal amplitude.

(d) $\eta_1 = \eta_2$ and of unrestricted magnitude

In this case

$$\left. \begin{aligned} \tan \psi_1 &= \tan \phi_1 = \sqrt{E}, \\ \tan \psi_2 &= \tan \phi_2 = 1/\sqrt{E}. \end{aligned} \right\} \dots (12)$$

Thus, when $\eta_2 = \eta_1(\epsilon_1\sigma_2 = \epsilon_2\sigma_1)$, no reflection is obtained with planes of equal phase coincident with planes of equal amplitude and at the Brewster angle.

(e) $\psi_1 = \psi_2 = 0$

As appears immediately from (7), (8), it is impossible for the reflected wave to be absent if $\psi_1 = \psi_2 = 0$, this is the case for the component (criss-cross) waves in hollow tubes, when the boundary between the two media is at right angles to the axis of the tube.

RECTANGULAR VOLTAGE WAVES FROM A LOW IMPEDANCE SOURCE

By T. J. REHFISCH,
Northampton Polytechnic
(Now at Electrical Research Association)

Demonstrated 27 October 1944

§ 1. INTRODUCTION

THE response of a physical system to a suddenly applied force has received much attention, electrical networks and Heaviside's Unit Function being of particular interest today.

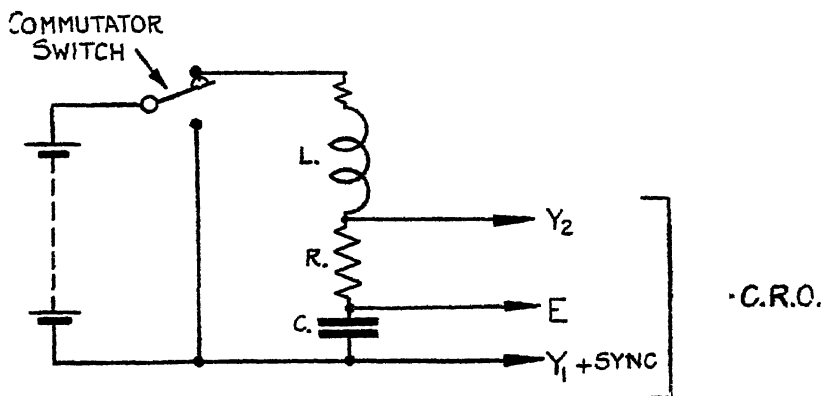
Experimental difficulties attaching to such "single transient" phenomena may, for some purposes, be overcome by periodic repetition of the transient, i.e. by the generation of a rectangular wave and its application to the test circuit; providing that the circuit reaches the steady state within the half-period of the wave, the transient behaviour of the circuit will be repeated over and over again and may be examined on an oscilloscope.

"Square-wave" generators using thermionic valves have been used extensively in recent times; however, there are reasons why a reversion to a mechanical "make-and-short" device is advantageous for demonstrations or even for work on scale models; primarily, because the internal impedance of this form of

generator is very small, being composed mainly of battery and brush-commutator contact resistances—a residual few ohms at most; this compares with many and unsteady thousands of ohms contributed by alternative low-power (thermionic) generators. Hence the mechanical device allows many circuit phenomena to be observed on a reduced frequency scale—in particular, oscillatory phenomena, which, with a high-impedance generator, occur in the radio-frequency range, may easily be obtained at moderate audio frequencies. Ordinary variable standard components—resistors, self- and mutual inductors, capacitors—may be readily used; their stray parameters, vitiating at radio frequencies, being of little importance in these experiments (with the exception of coil self-capacity), known and controllable changes in circuit parameters are readily achieved. Moreover, the low residual internal impedance of the mechanical generator may be increased at will by adding resistance externally; finally, by a simple re-connection, the “internal resistance” may be made different during the application of the p.d., and its removal, respectively. As an extreme case, a “make-and-break” arrangement leaves the circuit open whenever the battery is removed.

§ 2. APPARATUS

The generator consists of a battery feeding a motor-driven “Fleming-Clinton commutator”, modified by the addition of another centre brush; the two centre brushes are normally joined electrically and spaced accurately to ensure equal make-and-break periods and leave no observable insulating interval, without shortening the battery excessively.* The battery is connected between the outer brushes, and the output is taken from one of these and the centre pair.



Circuit for investigation of L - C - R (series) circuit

A commercial double-beam oscillograph is connected to the test circuits, one beam normally showing the p.d. wave-form existing across the condenser in the primary circuit, the second beam revealing the p.d. wave-form across a resistor in the primary circuit (current wave-form), or across a condenser in the secondary circuit. The time base of the C.R.O. is locked to the repetition rate of the generator (60 c.p.s.), or half this value, as the case may be.

* A paraffin lubricant was found essential for good contacts,

§3 OBSERVATIONS

Observations carried out with the equipment, some of which were demonstrated to the Society, are summarized in the following table, which shows the circuit and the parameters which can usefully be varied, with notes on the effects thus illustrated.

Circuit	Variable parameters	Observations of particular interest
A. Any	C R O time base B.F.O. output	Rectangular wave output unaffected by test circuit. Comparison with B F O sine-wave
B C-R in series	C, R	Condenser p.d. and its differential C/t , resistance p.d. (current). Rate of rise depends on $1/CR$; crest value of p.d.s unaffected by (CR) value, providing sufficient time is allowed
C. L-C-R in series (LC) small	L, C, R	Introduction of L speeds up attainment of steady state Critical damping. Verification of condition $\frac{1}{LC} - \frac{R^2}{4C^2} = 0$. Overshooting; damped oscillations of period determined by LC. Damping factor $R/2L$ independent of C
D L-C-R in series, (LC) large	C, R	Condenser p.d. overshoot limited to $< 2E$ When $2\pi\sqrt{LC} = T$ (repetition period), the circuit may be tuned to the applied wave (the fundamental and the second harmonic components) Condenser p.d. "purer" than current Coil p.d. very discontinuous
E. L-C- R_2 in parallel, R_1 in series Make and short Make and break	R_2 , L, C	Observations appear as duals of (D), but the current first "undershoots". Damping by R_2 leads to single-pulse p.d.; its width is shortened by decreasing L, C increasing R_2 ; its height limited by E. On open-circuiting the circuit wherein the coil carries a steady current, damped oscillations are produced; a 2-v. battery energizes it sufficiently to produce oscillations filling the screen (100-v. peak to peak). Their period depends on LC, their initial value on $\frac{1}{R_1} \sqrt{\frac{L}{C}}$, their damping on $1/F$ and C, but not on L At critical damping, a single pulse is generated.
Rectifier connected across the parallel circuit		This suppresses the oscillation completely or leaves the first half-cycle only, according to the sense of connexion

Circuit	Variable parameters	Observations of particular interest
F. Coupled circuits		
Both coils aperiodic	M	Make · rise of primary p.d. depends on M , non-oscillatory secondary current, direction determined by sign of M Short · oscillatory fall of p.d
Primary coil tuned	R_2	Damping of primary coil increased by reducing R_2
Both coils tuned	M	Circuits isochronous as M is increased, the damped oscillations in the primary circuit are broken up into separate packets, eventually too close to be distinguished; the envelopes of corresponding secondary circuit oscillations are in antiphase to the envelope of the primary.
		"Turn-over" points
	C_2	At 20% coupling, a maximum secondary oscillation occurs when the circuits are isochronous; smaller C_2 values effect substantial and more rapid secondary oscillations, while with C_2 some four or more times larger, the secondary condenser p.d. falls to practically zero, and the effect is similar to a short circuit on C_2 .
Secondary circuits also connected to source	M	Type of oscillation depends on mode of excitation—beats are suppressed, frequency determined by sign of M .

DISCUSSION

on papers by L. S. GODDARD (*Proc. Phys. Soc.*, **56**, 372 (1944)) and L. S. GODDARD and O. KLEMPERER (*ibid.* **56**, 378 (1944)) on Electron microscopes.

Dr L. J. COMRIE. I approve very much of this "back to nature" movement, in which the fundamental equations are treated rigorously, and I admire anyone who has the courage to face the necessary arithmetic. Perhaps the classical instance of this principle is Cowell's work on Halley's comet. Up to this time an elliptical orbit was used as a first approximation, and perturbations due to the major planets were applied. In this process, second-order terms, due to the difference between the assumed and the true position of the comet, eventually become burdensome. Cowell therefore discarded the elliptical orbit, and combined all attractions, *including that of the Sun*, in his step-by-step solution of the equations of motion. His success in predicting the return to perihelion to within two days over a period of 75 years (equivalent to 5-figure accuracy) is ample testimony to his treatment.

The W. E. Milne step-by-step method of numerical integration, as used by Mr. Goddard, is described in a series of pamphlets issued by the Marchant Calculating Machine Company, of which I am expecting a supply soon for distribution. My first examination of the method failed to convert me from the usual finite difference methods, especially those in which the computer estimates one or more higher-order differences of the highest differential and then produces (usually in pencil) enough central differences to use central-difference formulae to get values of the integral that suffice for use in the differential equation. The values of the differential obtained by substitution in the equation are

then compared with the estimated values. Although this is theoretically an iteration process, in practice a good computer will not need repetitions. I can speak from experience, as 14 of my girls, whose average age is less than 21, have just spent several months solving pairs of simultaneous differential equations of the second order.

The lack of attractiveness here is similar to that in the Lagrangian formula for interpolation and is due to (1) the loss of the sheet anchor of the differences, which not only detect errors, but also guide us to any necessary changes of interval; (2) the fact that the method is "blind" and cannot see how many function values to use, or detect errors in them easily. That such methods are more attractive with a machine like the latest electric Marchant is true, but I cannot resist the feeling that the method is here the slave of the machine, instead of the converse.

Mr. T. SMITH. I think the outstanding fact which emerges from these papers is that a method of calculation has been evolved which is quick, well-adapted to existing machines, and gives results in good agreement with experiment. From the description it appears that the method can be safely applied by computers who are by no means trained mathematicians. I am glad to see a method of this kind established, for it has seemed to me inappropriate to expect methods developed for glass optics* to be suitable for electron optical problems. The distinction between the abrupt changes characteristic of the one system and the continuous changes of the other appeared so important that I have not regarded very seriously any work on electron optics done by the usual optical methods. The one optical method which should undoubtedly be appropriate is that of the characteristic function, and I hope to discuss this some time with Mr. Goddard. As this method may be regarded indifferently as optical or dynamical—it is, in fact, the origin of Hamilton's equations of motion—there is no doubt that it has a strong *prima facie* claim to consideration.

Dr W. D. WRIGHT. Greater accuracy in the tracing of electron paths is obviously to be desired, and it would seem that Mr. Goddard's method represents an important advance over previous attempts. There may still, however, be a useful field of application for more approximate calculations, since the tracing of an electron through a particular lens is rarely undertaken because of any inherent interest in the path itself, but rather as a means to improving the design of a lens in respect of its aberrations. In the method adopted by Dr. Klemperer and myself, to which Mr. Goddard has referred, we sub-divided the system into a number of refracting surfaces and endeavoured to calculate the contributions which each surface made to the final aberration, by means of optical formulae developed by Conrady. When these contributions are known, it is possible to visualize the changes required in the refracting surfaces, that is, in the electrostatic field, in order to improve the aberrations. The fact that the data are only approximate is not really important, so long as the direction in which the aberrations will change with a given change in shape of the refracting surfaces can be deduced. I am not clear that the equivalent information can be obtained from an accurate method of ray tracing in which the field is treated as continuously variable, and it is certainly not possible to take advantage of the mass of information and experience accumulated in the design of glass lens systems. Our method was especially successful in providing a simple derivation of the Petzval curvature of field and in comparing the problem of flattening the field of an electron lens with that of designing an anastigmatic photographic lens.

AUTHORS' reply. Dr. Comrie's suggestion that a method using differences may be preferable to one using ordinates only is certainly worthy of consideration. As stated in the first paper, each method was tried and the method using ordinates proved to be much quicker, and so was adopted. But this was the sole reason, and we agree that the method using differences is safer, and in some problems it may actually be more rapid. If, for example, the conditions of the field or the electron change a great deal over the portion

* It is often forgotten that the current trigonometric methods of lens design were adopted, not because they were, without qualification, the best for the purpose, but because, taking into consideration tabular and other aids to calculation available at the time, they were considered the methods most suitable for application by workers merely trained as routine computers. Apart from vested interests of one kind and another, I think they have little claim to serious consideration today. What is important is the accumulated knowledge that has been gained by their use, and this in different circumstances would have been obtained in other ways.

of the path to be determined, then the problem of a change of interval may arise. In this case the method recommended by Dr. Comrie might well prove to be superior because it would reveal when the interval should be altered and what alteration is necessary. This problem, however, did not arise in the case of the magnetic lens.

Mr Smith's remark about the possibility of using the method of the characteristic function is very interesting. It is also very valuable because this method, when it can be applied in practice, is one of great power, and there has been a tendency in some quarters to neglect its possibilities. The refractive index, in the electron-optical case, is given by

$$\mu(r, \mathfrak{s}) = \phi^{\frac{1}{2}} - \eta(\mathbf{A}\mathfrak{s}) = \phi^{\frac{1}{2}} + \left(\frac{e}{2m}\right)^{\frac{1}{2}} (\mathbf{A}\mathfrak{s}),$$

where ϕ and \mathbf{A} are the electric and magnetic potentials respectively, and \mathfrak{s} is the vector,

$$\mathfrak{s} \equiv \left(\frac{dr}{d\sigma}, \frac{d\psi}{d\sigma}, \frac{dz}{d\sigma} \right),$$

obtained by differentiating along the path. From this it is evident that the medium is heterogeneous and, when a magnetic field is present, also anisotropic. This shows how inadequate it is to expect much help from the analogy of glass optics where the medium is homogeneous and isotropic. It should be stated that the method of the characteristic function, or at least an equivalent, has been studied by Glaser in a series of papers in the *Zeitschrift für Physik*. However, even Glaser does not appear to have considered the full possibilities of the method. A further study is certainly desirable, as Mr. Smith suggests.

We agree with Dr. Wright that there is seldom any intrinsic interest in the path itself, and this fact was appreciated at the time the papers were written. The problem was not so much to develop a method for very accurate ray tracing as to provide a method which was general and at the same time did not introduce the very large errors that are inherent in some of the methods used hitherto. In the case of magnetic lenses, ray-tracing proved to be impracticable for reasons pointed out in the second paper, and it was necessary to develop an alternative method. It was possible to solve this problem only by freeing oneself from the methods of ray tracing which had become standard during the last ten years or so. Only by a return to the exact equations of motion and a numerical solution of these by a method that has a rigorous mathematical derivation could one hope to achieve the desired accuracy and generality. The new method, considering the accuracy of its results, is considerably quicker (when used in conjunction with a modern calculating machine) than the methods hitherto developed. No time is lost in referring to tables of functions or in plotting or using graphs, and it is unnecessary to calculate any auxiliary quantities. The method is, in fact, an *algebraic* process.

In the case of trigonometric ray tracing, the convergence is slow, as may be seen by referring to table 1 on p. 300 of the paper by Dr. Klemperer and Dr. Wright (*Proc. Phys. Soc.* **51**, 296 (1939)). By solving numerically the equations of motion it would be possible, for a given accuracy, to use a much smaller number of points because, in passing from a point P_n to P_{n+1} , use is made of much more information than merely the conditions existing at P_n .

Regarding the estimation of lens errors, it should be noted that these may be calculated by means of formulae developed in the series of papers by Glaser, mentioned above; and these may well prove to be of importance in the design of electron-optical systems. For this purpose it seems doubtful whether extensive use could be made of experience gained in glass optics, and on this point it is worth noting the remarks made by Mr. Smith. In glass optics the refractive index of a medium may be changed without altering that of neighbouring media, but in electron optics the system must be considered *as a whole*, and the reduction of errors will be achieved through a consideration of integral rather than differential (or local) conditions.

For the Petzval curvature an exact expression has been given in terms of the field strength (see, for example, W. Glaser, *Z. Phys.* **97**, 177 (1935), and, in particular, p. 195). We have, using Glaser's notation, $C-D$ for the Petzval curvature,

$$C-D = \frac{1}{8} \int_a^b \frac{\phi''(x) + 4\eta^2 H^2(x)}{\{\phi(x)\}^{\frac{3}{2}}} dx,$$

where $\phi(z)$ and $H(z)$ are the scalar potential and magnetic field strength along the axis. In the case of a purely electrostatic lens we have $H(z) \equiv 0$, so that

$$\phi''(z)$$

This formula* gives the curvature more rapidly than that used by O. Klemperer and W. D. Wright, as only a single (numerical) integration is needed. It is also evident that the contribution of different parts of the field to the curvature may be found by studying the function $\phi''(z)/\{\phi(z)\}^{\frac{1}{2}}$. It is hoped to treat the question of aberrations of electrostatic and magnetic lenses in a future paper.

* The formula used by Dr. Wright may be easily derived from that given. The refractive index μ is given by $\mu = k\phi^{\frac{1}{2}}$, where k is constant, and the radius of curvature of the equipotentials (on the axis) is (see Myers, *Electron Optics* (London, 1939), p. 95)

$$R = 2\phi'(z)/\phi''(z).$$

Hence

$$C - D = \frac{1}{2} \int_a^b \frac{\phi'(z)}{R_1 \phi(z)^{\frac{1}{2}}} dz = \frac{1}{2} k \int_a^b \frac{\mu' dz}{R \mu^{\frac{3}{2}}}.$$

If the integral is now replaced approximately by a sum we obtain the desired form

CORRIGENDUM

Paper on "A mean scotopic visibility curve, by W. S. STILES and T. SMITH
(*Proc. Phys. Soc.* 54, 251 (part 4, 1944))

The measurements of Hecht and Williams (1922) were made for binocular vision, *not* monocular vision as stated, and the number of observers was 48. The authors are indebted to Prof. Hecht for this additional information.

RECENT REPORTS AND CATALOGUES

Copper and Copper Alloy Springs (C.D.A. Publication no. 39, 1944) Pp. 62. THE COPPER DEVELOPMENT ASSOCIATION, 9 Bilton Road, Rugby, Warwickshire.

Annotated Catalogue of Works on Physics, including also items on collateral sciences, and comprising the Library of John Tyndall, F.R.S. (1820-93), with a supplement of periodical publications and addenda. (Catalogue no. 873, 1944.) Pp. 88. HENRY SOTHERAN, LTD., 2-5 Sackville Street, Piccadilly, London W. 1.

Antiscatter Treatments for Glass, by F. W. REINHART, RUTH A. KRONSTADT and G. M. KLINE (National Bureau of Standards Miscellaneous Publication M 175, June 1944.) Pp. 31. U.S. DEPARTMENT OF COMMERCE, Washington, D.C. 10 cents.

Discussion of the National Electrical Safety Code Part 2 and Grounding Rules. (National Bureau of Standards Handbook H 39, July 1944.) Pp. 162. U.S. DEPARTMENT OF COMMERCE, Washington, D.C. 75 cents.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 57, PART 2

1 March 1945

No. 320

THE FORMATION OF METAL-SPRAYED DEPOSITS

By W. E. BALLARD,
Dudley

MS. received 25 October 1944

ABSTRACT. During the last two decades many papers on metal spraying have appeared, some of which are of very great technical importance, though very little of the published work has given any indication of the fundamentals of the process.

The difficulties of research on the process are examined, these difficulties being increased by the rapidity of the cycle of events in the process of wire spraying. Some difficulties have been overcome by the use of the high-speed cine camera, which has indicated that the spray of metal is rapidly pulsating and that a retraction of the deposited agglomerates of particles takes place on the surface. A theory is put forward to show that surface tension plays a considerable role both in the pulsation of the spray and in the ultimate structure of the coating. Some indications are given that the amount of metal sprayed in unit time is controlled by well-known physical laws, and an empirical formula for speed of working is given and examined. Developing the theory of the action of surface tension at the moment of impact, explanations are given for certain phenomena observed in depositing the metal on smooth glass and roughened metallic surfaces, including the problem of adhesion. Some effects due to the spread of the spray, marginal deposits and the temperature of deposition are noted, and a tentative explanation for the varying percentage loss with different metals is advanced.

§1 INTRODUCTION

SINCE the invention of the metal-spraying processes by Schoop, in 1910, there has been published a great mass of technical literature; unfortunately a large percentage of this has described in detail special makes of tools and has been coloured by trade influence.

In the *Journal of the Institute of Metals* there may be found three papers describing metal spraying in various stages of development. The first, by Morcom (1924), was an account of what was then an entirely novel conception in the art of working metal. In 1924 the process was developing commercially in this country, and T. H. Turner and the present author described the wire process and the properties of the deposited metal. A period of considerable expansion in the industry then took place, and the third paper was contributed in 1937 by E. C. Rollason, in which the author surveyed a somewhat wider field and compared the results obtained by the three spraying processes in use in commerce. Since that date, the use of sprayed metal has become general in many branches of engineering and very great advances have been made, both in the apparatus used and in the application of the deposits. The present paper is not intended to describe these advances, interesting though they may be, but rather to endeavour to explore some of the fundamental physical considerations

involved in the art of metal spraying. The paper deals specifically with the wire process unless otherwise noted, but the comments do not refer to any one type of apparatus.

Comparatively few papers in the literature of the subject have described work of a fundamental character, and there are many contradictions and ambiguities. This is undoubtedly due to the many variables which affect the process, and the very great difficulties presented in exact evaluation. For example, it might appear to be easy to measure the exact adhesion of sprayed metal to the base by one of the methods developed for electro-deposits (see Hothersall, 1943); but in fact such methods are not applicable because it is necessary that the deposit be built up to sufficient thickness to resist the test load, and this increase of thickness will, unless the shape of the test piece be very carefully chosen, seriously affect the result. Methods relying on soldering the sprayed coating to the apparatus must also be suspect because even slight heating affects internal stress.

Recently Ingham and Wilson (1944) have described a method of testing adhesion which seeks to overcome these difficulties by depositing the metal on a shaft in which there is inserted a close-fitting plug. The method would seem to be very promising, but even so it is not entirely certain that the results give the measure of true adhesion. This does not mean that adhesion tests are valueless; it merely indicates difficulties in exploring the problems of spraying in what would appear to be a straightforward example.

Schoop (see Rollason, 1939) believed that the metal particles shot from the gun became solid during transit and became molten again at the moment of impact, due to the release of the kinetic energy. This view-point was contested by Arnold (1917) in one of the most instructive papers on the subject that has been written. Arnold's calculations are, however, open to question because he assumes that the temperature of each metal particle at the moment of impact is that of the surrounding envelope of gas. This is extremely unlikely. Furthermore, the temperature of 70° c. assumed by Arnold for the gaseous envelope is far too low, as will be seen later.

Turner and Ballard (1924) and Rollason (1937) have shown photographs of splashes of sprayed metal on glass, indicating that the metal is completely liquid in some cases, and Rollason, by means of back-reflection x-ray photographs, has shown evidence of some cold work in the deposits.

All workers agree that there is no alloying of the sprayed metal to the base metal, but many theories have been advanced to account for the adhesion of the deposit, none wholly satisfactory. The simplest idea appears to be that the deposit interlocks with the interstices of the shot-blasted base, giving a type of microscopical dove-tailing. This is undoubtedly true, but a molten metal allowed to solidify on a shot-blasted mould will also fill these interstices, though the adhesion will not be as high as that of a sprayed layer. There would seem, therefore, to be some other factor involved. In the paper of 1924, Turner and the author wrote: "It is possible that these illustrations may prove of interest to those engaged upon the problem of surface tension in metals". In this paper an attempt will be made to show that this suggestion contained a possible explanation of some of the phenomena of metal spraying.

Before considering this further, it should be appreciated that the difficulties of investigating these problems are intensified by the rapidity of the events of the process. The metal is melted, is divided into small particles, travels through the spraying distance and is deposited, all in an extremely small part of one second. It might be argued that a similar sequence of events, perhaps of a more complicated nature, takes place in electro-deposition, that is, the tearing off of metal from the anode, passing into chemical combination, migration, decomposition and deposition at the cathode, all of which may take place, under some conditions, in a short time. It should be remembered that although the particle size in spraying is very small, it is certainly very large when compared with the ion in electrolysis, and, therefore, the processes are not comparable. Except for Arnold's work, there are no definite data available on the speed of the sprayed particles.

It will be admitted that for a metal to be sprayed from a wire it must be melted. It is possible, by speeding up the wire-feed to the flame, to get lumps of plastic metal torn from the wire, but they are immediately apparent, being large and irregular in shape and not being a normal feature of the process. The first assumption is, therefore, not affected by this possibility.

Now the surface tension of metals is known to be very high (see Burden, 1940), and therefore as each drop of metal is pulled from the wire by the surrounding gases, it will assume a spherical shape, and it is reasonable to suppose that each drop will be accompanied by a very tiny drop, the well-known Plateau spherule. The author believes that much of the exceedingly fine dust generated during spraying is caused by the dissemination in the atmosphere of these spherules. They are so small that they cool quickly, and unless entrapped by a larger particle do not adhere to the surface to be sprayed but are carried away by the gases. Occasionally, however, one is so entrapped and can be seen in the micro-structure of the sprayed metal. Figure 1 shows these bodies in a sample of sprayed medium carbon steel. In the case of zinc, the existence of these spherules in the air around a pistol probably accounts for the beneficial physiological effects on the operators observed and reported elsewhere by the author (1943). The existence of the spherules does not modify the structure of the deposits to any extent, and their existence is merely noted as a matter of interest.

The spherical drop of metal, having left the wire, commences on its journey towards its final resting place in the deposit, and by the time this is reached, the drop

(i) may still be liquid;

(N.B. Or it may be solid, and so hot that the release of kinetic energy on impact remelts it, in which case it may be regarded as liquid for the purpose of this argument.)

(ii) it may be solidified, but hot enough to be plastic;

(iii) it may be solid, cool and not easily deformed.

Taking the last case first, because it is the least complicated, it will be noticed that the velocity of the particles is comparatively high, and it would be normal

to expect that the coefficient of restitution between the particle and the base would also be comparatively high. In other words, one would expect such a particle to bounce and be carried away by the gas stream, except in the case in which, by collision with another particle, it becomes entrapped by aggregation. Examination of the micro-structure of sprayed metal confirms this view, as entrapped bodies, roughly spherical in form, are not a general feature of the structure, but they occasionally occur at great nozzle distances, when it would be expected that more of the particles would have cooled and have lost their initial high velocity. Such an entrapment was shown by Rollason (1937), and another such body in sprayed aluminium is shown in figure 2.

Although these entrapments are unusual in wire-sprayed coatings, they are common in powder-sprayed coatings, because some powder always escapes the hot zone of the flame and remains solid, but becomes mixed with more plastic elements.

The behaviour of the spherical drops which are liquid at the time of impact is more interesting. If a light layer of sprayed metal is formed on glass it can be examined by transmitted light, and is found to consist of particles which have a definite splash formation. Figure 3 shows such a splash of sprayed copper. This formation can best be understood by reference to the work of Professor A. M. Worthington (1894), who studied the behaviour of drops of liquid falling on to a surface which they did not wet. The patterns of the splashes in metal-sprayed deposits are often similar in design to the formations observed towards the end of the cycle of incidents before the drop becomes static, and this indicates, therefore, that the freezing of the particles is not instantaneous. The possible results of these effects will be discussed later in the paper.

The remaining class of drops includes those which, although solid, are hot enough to remain plastic, and these probably represent a fairly large proportion of the marginal spray. On impact, their spherical form will be considerably flattened, particularly at the base, that is, the area of impact and the top will be flattened by the impact of the next particle arriving. The final shape therefore is that of a plate. These particles are not seriously deformed at the top on the outside of a deposit, as they have not received blows from above, and, therefore, they are the reason for the matt appearance of the outer surface of a deposit.

In practice, spraying is often carried out in layers, that is, a surface is coated with a light deposit and the pistol then traverses the same area again to give subsequent layers. The particles on the top of each layer will be less deformed, and, therefore, the micro-structure of a coating formed in this way will show broader bands of undeformed particles. Figures 4 and 5 show respectively a coating of copper built up continuously and a similar coating built up in two passes, and in the latter the broad band may be noted.

This type of particle is, of course, subject to some cold work, and the internal structure is equi-axed, whereas the structure of the interior of the splashed drops is columnar. Examples of these two types of structure were given in Rollason's paper.

Having now briefly described the types of units one expects in sprayed coatings, it is possible to examine some of the factors influencing their distribution and some of the results of their presence.

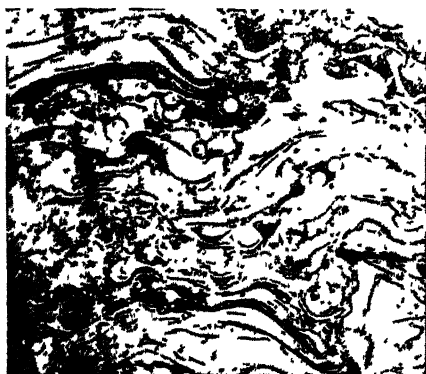


Figure 1 0.4% carbon steel as sprayed, showing spherules $\times 200$



Figure 2 Aluminium as sprayed, showing entrapped particle $\times 200$

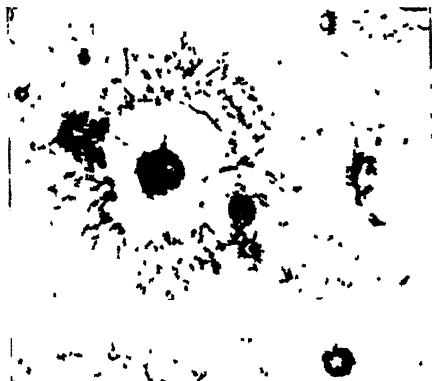


Figure 3 Copper particle sprayed on cold glass $\times 75$.



Figure 4 Copper sprayed continuously. $\times 200$

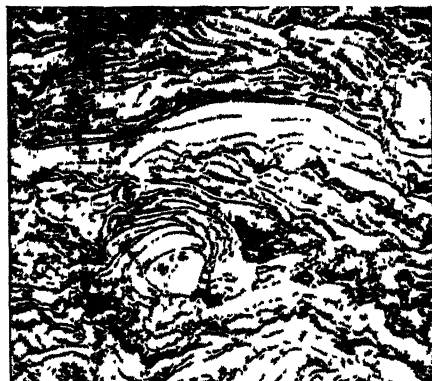


Figure 5. Copper sprayed in layers, showing one top layer $\times 200$.



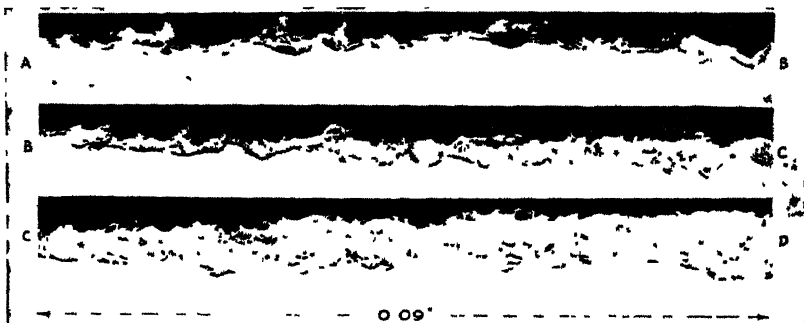
Figure 7. Ends of zinc and aluminium wire. $\times 6$.



Figure 8. Still from film showing nozzle and spray About actual size



Figure 9 As figure 8, but about 1/300th of a second afterwards



Transverse section of deposit of 0.15% C steel from one pass of the pistol. Base shotblasted-steel

Figure 11 Composite micrographic section of steel deposit



Figure 14. Under-surface of tin sprayed on glass $\times 30$.

§ 2 TEMPERATURE OF GAS STREAM

Arnold assumed that the temperature of the gas stream at the working distance indicated the probable temperature of the metal particles within the stream. This was disproved by Thorman (1933), who, by means of a radiation pyrometer, sought to show that iron particles in the spray were at temperatures exceeding 1000°C . It may be suggested that this method is questionable because the majority of the particles which give incandescence to the iron spray may be themselves burning in the oxygen in the air blast. This is unlikely, as it cannot be disputed that at moderate working distances iron shows a splash effect. Furthermore, oxidation is not predominant, as the metal magnesium sprays with ease and shows little sign of burning.

It seems reasonable to suppose that the hotter the gas stream at the spraying distance the hotter will be the particles in general and the greater will be the

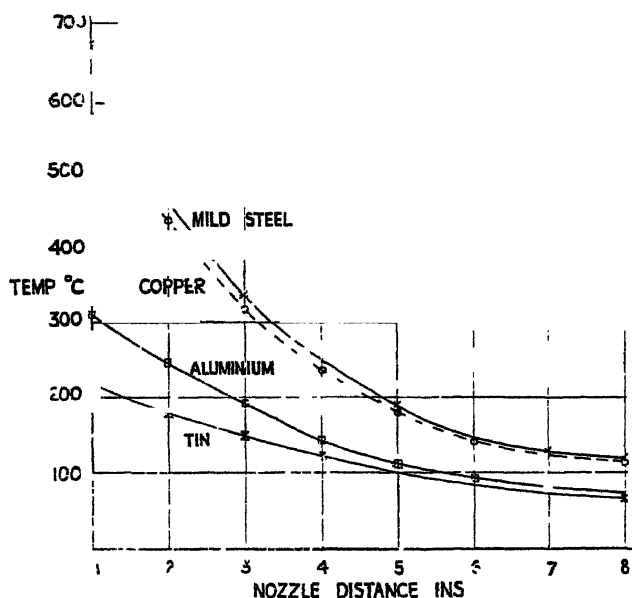


Figure 6 Curves showing temperature of gas stream at varying nozzle distances.

proportion of liquid drops. The temperature of the gases will of course depend largely on the type of pistol and the amount of combustible gas, oxygen and air that is used, besides bearing a relationship to the amount of metal sprayed and its melting point. Nevertheless, it seemed desirable to investigate the usual range of temperatures. This was done by exposing metal thermocouples of very fine-gauge nickel and nichrome wire, carefully standardized, immediately before the nozzle at measured distances. Careful blank tests were carried out to establish any effect of the metal deposit sticking to the wires. The amount of metal adhering was very small and it did not make any difference to the readings. The results are shown in figure 6, the pistol being of a standard pattern, the fuel compressed coal gas, and the wire 2 mm. diameter. It is apparent at once that the spraying process is not as cold as some have claimed. The form of the curve is interesting as it approximates quite closely to the logarithmic curve

expressing the law for bodies cooling in an air stream. It would be expected that greater divergence would be shown, because heat is absorbed by the expansion of the compressed air stream and by the air from the room, which is drawn into the pistol blast in very large quantities. It should be noted that in these experiments the thermocouple is placed in the axis of the wire, and therefore represents the maximum temperature of the gas stream at that particular nozzle distance, the outer streams of gas being much cooler. It is interesting also to note that the curve showing the pressure of the gas stream at varying nozzle distances is exactly similar in form to that showing temperature drop.

That the temperature of the gas stream is comparatively high is illustrated by the following experiment. A steel tube, 2" long \times 0.6" o.d., was held in the chuck of a lathe, and a thermocouple was placed inside the tube. Brass was then sprayed on the tube with the pistol at varying nozzle distances. The time taken to reach a constant maximum temperature was noted, as was also the temperature in $^{\circ}\text{C}$.

Nozzle distance (in.)	Time to reach maximum (min.)	Maximum temperature ($^{\circ}\text{C}$.)
1	4.5	360
2	6	280
3	7	165
4	7.5	145
5	8	125
6	8.5	120
7	10	105
8	11.5	95

It is apparent from these figures that even if it is considered likely that the temperature of the particles and the gaseous stream are equal, Arnold's calculations showing that it is impossible to suppose that the particles reaching the surface could not be remelted by the release of kinetic energy are based on error, as he assumed a temperature on arrival of 70°C .

§ 3. THE FORMATION OF THE DROPS AT THE END OF THE WIRE

It was noticed long ago that if the flame of the pistol is extinguished suddenly by pinching the oxygen pipe, the wire will still be advanced by the pistol gearing, and the end of the wire which has been in the flame will show a conical form. This is only to be expected, and one would also assume that the cone will be longer the thicker the wire and the lower its conductivity. This assumption is found to be true by observation, the cone of 2-mm. lead wire having a smaller angle included at its apex than a 1.5-mm. copper wire. Further information can be gained by observation of these wire ends. Metals of higher melting point, e.g. nickel, often show spirally arranged grooves on the sides of the cone. Many metals show a thickening at the base of the cone, that is, the diameter of the base of the cone is slightly larger than the diameter of the original wire. This effect is particularly noticeable in the case of aluminium or aluminium alloys, see figure 7.

In order to examine more fully the spiral formation and the reason for this thickening, use was made of the high-speed cinematograph camera. The work was undertaken with the help of the research staff of Messrs. Kodak Limited. The camera exposes about 100 ft. of 16-mm. film in one second, and in one second between 2700 and 3000 pictures are taken. The 100-ft. length of film takes about five minutes to pass through the projector. The speed of the subject taken is therefore reduced about 300 times. Some stills from the film are shown in figures 8 and 9, which depict the spraying of a zinc wire, 2 mm. diameter, using propane as the combustible gas. It is unfortunate that while such a film is extremely revealing on projection, stills such as these reproduced from it are somewhat unsatisfactory, due in part to the texture of the sensitized layer. If the wire be examined after the pistol is stopped, zinc does not usually show the spiral effect, but the film shows that the spray is actually spiral in form, the helix being very open. The film shows also that the spray is not quite so steady as it appears to the ordinary observer, but pulsates rapidly, the periodicity being of the order of one three-hundredth of a second. The metal appears to melt on the side of the cone and to be sucked into the spiral of the flame zone. Then the suction becomes less, or surface tension increases, and the thin film of molten metals runs back and collects as a ridge on the base of the cone. Then suction again predominates and the cycle of events recommences. Whether the action of the thin film of metal on the cone side gives rise to the twisting action, or (as seems more likely) the expanding gases tend to take a swirling path, cannot be stated with certainty at the moment, but it is hoped to carry out further experiments by means of a different type of camera and illumination in the future. The pulsations may be due to irregularities in the wire feed, but this does not seem likely, as they are quite regular. The recession of the molten film towards the base due to surface tension would in itself produce the rapid regular pulsations observed. The particles themselves are too small to be examined by the ciné method, but work will be continued by means of spark photographs.

If it be accepted that the explanation of the phenomenon occurring at the end of the wire is as described, then it should be possible to make further deductions on this basis. The metal-spraying pistol will work equally well under water, and it is, therefore, a comparatively simple matter to determine moderately accurately the total heat generated in the process. All that is necessary is that the pistol be held for a certain time in an insulated vessel containing water, the temperature of which is known, and the rise in temperature of the water can be noted and the heat calculated, making allowances for the heat content of the pistol itself, and carefully checking the temperature of the air coming away from the surface of the water. By carefully arranging the apparatus, results have been obtained which are extremely close to the theoretical values obtained by taking note of the volume and the calorific value of the fuel gas. The amount of heat required to melt the metal passing through the pistol can also be calculated within reasonable limits, and a number of experiments under various conditions have indicated that the maximum amount of heat absorbed by the metal itself is seldom more than 5% of the total heat generated in the flame. The major part of the heat is dissipated in raising the temperature of the air which is sucked in from the surrounding atmosphere by the injector effect of the burning gases

and of the propellant leaving the nozzle at high velocity. In these circumstances it is correct to assume that the metal entering the flame reacts as a cold body entering an atmosphere in which there is a huge excess of available heat. Therefore it follows that the temperature of the flame is high compared with that of the wire, and there will be a tendency for the film of metal on the outside of the wire to gather heat extremely quickly and become molten. As soon as it is molten it will be torn away by the suction effect. Where a metal has a very high conductivity, as is the case with copper and aluminium, the heat will pass very quickly from the surface film towards the centre of the wire, and the film of molten metal will tend to be thicker than the film of a metal with a lower conductivity. This is probably the explanation of the fact that the particle size of copper and aluminium tends to be larger than that of zinc or iron when sprayed under the best conditions.

Proceeding with the argument, it follows that if melting occurs in films as stated, the area of the hot film will have a predominating influence. Rejecting for the moment the length of the cone at the end of the wire, this area, which is always being renewed, will be dependent on the volume of wire used in a given interval of time, and this again, if the diameter of the wire is the same, will depend on the speed of wire passing. If the diameter of the wire be altered, then the area will also be subject to change, but if this is correct, then a change in wire diameter will bring about a proportional change in the volume of wire passing. There may, at first sight, appear to be an objection to this argument because the amount of heat taken up by various metals will be different in accordance with their specific heats and their latent heats of fusion. If, however, there is a large excess of heat available, then this factor would not appear to be so important as the area of metal presented to the flame, this area being, as already stated, largely dependent on wire diameter. The other factor which will affect the area considerably is the length of the cone if it varies within very wide limits, but as a rule the length of the cone is not very widely different with the same diameter of wire, although, as already stated, there is some variation.

If the length of wire sprayed in unit time be plotted against the melting point of the metal, the curve will, under certain conditions, be logarithmic, and it is found that with metals melting under 700° c. it will take the form which satisfies the equation

$$T = Ke^{-l},$$

where T = temperature (°c.), K = a constant, l = length of wire sprayed in unit time (ft. per min.).

This is closely parallel to the law for bodies cooling or being heated under these conditions. The value of K can be found by experiment for any one diameter of wire and spraying condition, but the reason for any particular value does not appear to be clear. Actually, the hot gases are in the state of turbulent flow, and the pick-up of heat will be dependent on transfer through the interfaces, i.e. hot gas to molten film and molten film to solid metal.

The value of K will vary to some extent with the particular fuel gas used and the shape of the nozzle. It is probably a composite of several factors, including dimensionless groups similar to the Reynolds number. The curves shown in figure 10 apply to a multi-fluted nozzle, using propane as the fuel gas. If the

same gas is used throughout and the nozzles are similar in construction, K is directly proportional to wire diameter. In the curves shown, $K=2100$ for wire of 1.5 mm. diameter. These facts seem to add confirmation to the theory of film formation on the wire tip.

It has been mentioned that this is only true for those metals or alloys melting at temperatures below 700°C . If the curves are extended above this range, they indicate false values in the case of the larger-diameter wire. The higher curve for 1.5-mm. wire appears to be merely an extension of the curve below 700°C ., i.e. the value of K remains unchanged. For 2-mm. wire the value of K is considerably reduced, and for 11-gauge B.S. the value of K closely approximates to that for 2-mm. wire below 700°C . This indicates that, considered from the point of view of thermal efficiency, the spraying of thick wires of the higher

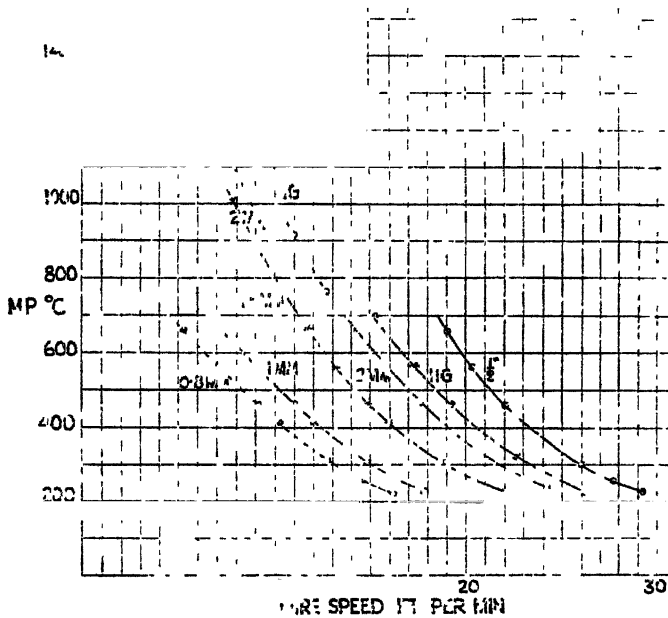


Figure 10. Curve showing relationship between melting point and wire speed.

melting-point range is not desirable. In practice this is recognized, e.g. in American practice, where the tendency is to use $\frac{1}{8}$ " diameter wire for the lower range and 11 B.S. for the high range. The reason for this can be indicated. If one observes the end of the wire while spraying metals in the high range, it is seen that the wire tip is visibly hot, and as radiation increases rapidly with temperature one would expect a high heat-loss by radiation from the cone surface. This will reduce considerably the value of K . It is curious, however, that the curves as found by experiment take the same form as those in the lower range, at least for melting points up to 1500°C ., because as the amount of radiation increases rapidly with temperature it might be reasoned that the slope should be greater. Furthermore, the usual form of such curves requires the use of another constant in the negative power of the exponential, but in this case its

value is undetermined. Therefore these curves and formula can only be considered as empirical, but giving some indication of the mode of formation of the spray.

Another point with regard to the curves, showing relationships of the wire speed to temperature, is that it is interesting to note that the curves for very fine gauges of wire are misleading, as it is possible to spray at a very much faster rate than the curves indicate. The curves cross the temperature axis, which means that it should be impossible to spray small-diameter wires having a melting point above that shown by the intersection. This is certainly due to the fact that the reservoir of heat in the flame is so great that, with the thin wires, melting takes place by the formation of a small drop at the end of the wire rather than by the tearing of a film which has been described in the more commercial sizes of wire. Actually, this drop formation and the disintegration of the drop by the propellant gas is quite a satisfactory method of spraying, and, in fact, the coatings produced are much finer in texture because the metal has excess of heat, and is consequently broken down into fine sprayed particles much more easily by the gas stream. Hence in practice very fine coatings of steel can be obtained with 0.8 mm. or 1 mm. diameter wire.

§ 4. THE PARTICLES IN FLIGHT

Little information is as yet available regarding the speed of transit. Arnold in 1917 carried out experiments to determine the speed of sprayed particles, and concluded that it approximated to 130 metres per second at normal spraying distances. There is no reason to dispute Arnold's work, but it is hoped to check it in the near future by means of high-speed photography. The actual speed, however, is not of great importance in considering the final coating. There is no reason to suppose that during transit the particles are other than spherical in shape, and it is extremely likely that they have a spinning motion in view of the spiral form of the mass of the spray.

§ 5 THE FORM OF THE SPRAY

It would be expected that in common with the usual types of spray the metalliferous spray would be concentrated in the central zone and more diffuse in the outer portions. That this is so is apparent if one watches the spray of steel, this spray being itself luminous. What is not generally realized is that the central zone is comparatively narrow and the metal content highly concentrated. This can easily be shown by drawing the pistol along a suitably roughened surface at normal spraying distance and examining the result.

Figure 11 shows a composite microphotograph of a section of the deposit of steel from one pass of the pistol passing and depositing on to a shot-blasted surface of steel. It will be seen how loosely packed are the particles in the marginal deposit, and also how loosely it adheres to the surface until covered with the subsequent layers. It will be realized therefore that the porosity of sprayed coatings is largely dependent on the marginal particles. This explains the increase in porosity with increasing nozzle distances. It may be assumed that it might be preferable to spray very heavy layers and gradually cover the surface by slow movement, but this is not true because the marginal particles cause

porous layers to form on each side of the spraying area, and coarse porosity occurs from the base metal to the surface. This is often more dangerous than porous layers lying parallel to the base.

Reference to figure 12 shows the actual zones diagrammatically. The central zone responsible for the greater part of the deposit lies between AA in the case of steel. The spray is luminous between BB and some small deposit occurs in this area. Metals of lower melting point, e.g. zinc, give sprays containing much more metal (i.e. the metal consumption and deposition are greater), and these sprays tend to have a slightly wider throw, the particular pistol used spreading

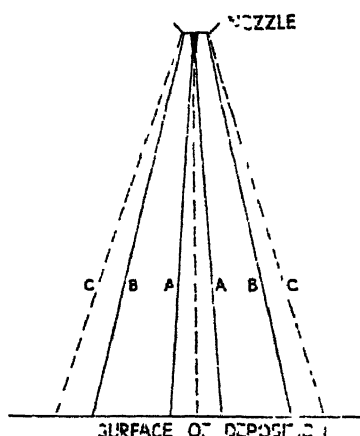


Figure 12. Diagram of form of spray

between the limits CC. It is often suggested that pistols having a large spread should be made, i.e. "that a white-wash brush is needed instead of a camel-hair pencil". It will be seen that all the metal must come from a focal point, i.e. the apex of the wire cone, and increase in wire diameter does not necessarily mean larger spread. There are methods of increasing the area of deposit (e.g. fan-nozzles), but these necessitate many particles travelling great distances, and possibly becoming cooler. For that which follows it will be seen that this may not be advantageous.

§ 6 THE DEPOSITION OF THE PARTICLES

In considering the events at the time of impact, it is desirable to consider deposition:—

1. On smooth surfaces, such as glass. (This, of course, is never carried out commercially.)
2. On suitably roughened surfaces, usually of metal.

(a) Deposition on glass

The surface of glass can be considered for the purpose of this discussion as being smooth, and it is also a surface which is not wetted by liquid metals, for example mercury. At the same time it is admitted that at exceedingly high temperatures, above the softening point of glass, some effect similar to wetting might be noticed. Such effects do not appear to be seen in metal spraying.

Also, glass is a poor conductor of heat (about 0.0025 calories $\text{cm}^{-1} \text{sec}^{-1}$, compared with 0.167 for iron and 1.041 for copper). That surface conductivity might have some effect must be realized, although the major dissipation of heat is caused by the gas stream. This was pointed out by Parkes in a written communication to Rollason's paper (1937). Also, by reference to the experiments already described, it will be seen that the steel tube sprayed with brass at 1" distance reached a maximum of 360° C. The temperature of the gas stream was 510° C. at this distance. It follows, therefore, that in this case some dissipation does occur by way of the receiving surface, and also, that a hot particle arriving at a glass surface will cool slightly less quickly than a similar particle arriving on a steel surface.

Reference to the work of Turner and Rollason shows that the particles first deposited on glass surfaces form well defined splash figures, as shown in figure 3.

Now Worthington (1894) showed that drops of mercury falling on glass assume well defined forms, some of which are shown diagrammatically in figure 13. It is possible, also, by careful visual examination of spots of water on clean glass, to follow similar events. It is, of course, realized that here we are only concerned with those particles which arrive at the surface as molten drops or become molten at the moment of impact.



Figure 13. Diagram showing splash forms.

A molten drop, arriving at the surface which it does not wet, will first spread out as a flat plate, and will, if surface tension does not operate, continue to spread until the plate became of unimolecular thickness or the drop solidifies. The spreading is of course helped if the drop has arrived at high velocity, but even then, after a short interval, surface tension restrains this spreading tendency, the edge of the plate becomes thickened and there is a tendency for the thickened edge to break and form a ring of small spherical drops, leaving a spherical drop in the centre, smaller than the original. This action of retraction also causes a number of radial splashes. In the case of the sprayed drop, solidification may occur at any instant in the cycle of events. After examining the sprayed particles on glass, it will be seen that most of the forms of splashes described by Worthington are present. It is apparent, therefore, that in considering the formation of sprayed metal coatings, retraction of individual drops or aggregates of drops, due to surface tension, must be taken into account. Oxide films on the particles or aggregates also cause modifications to the splash effect, as they tend to interfere with the distribution of surface energy. A film of the arrival of drops of zinc and of steel on glass was prepared by the high-speed camera. The camera was focused on a glass plate and the plate was sprayed from the other side. On showing the film at normal speeds, the retraction of some of the

agglomerates of particles can be observed. It is unfortunate that it has been found impossible to take from the film representative stills showing this effect. In the case of some metals, most of the particles are frozen in the first stages of spreading and retraction and most of the particles appear roughly circular in shape, with serrated edges. This is shown by cadmium, lead, tin and zinc, and these metals have a comparatively low heat-content when fused. On the other hand, aluminium, iron, nickel and copper have high heat-content, and these metals show more variation in the pattern of the splashes.

One aspect of this is perhaps shown as follows. Consider an ideal case, never obtained in practice, of all the particles being molten on impact. All particles of the metals with high heat-content will give pronounced splash effects and the patterns will be much broken up. Surface tension will have caused many small spheres, not connected if the layer be one particle in thickness. If the metal has low heat-content, the drops will be frozen when the surface tension is at its maximum value, considering the flattened drop as a whole. These metals will show, in general, higher adhesion, for reasons discussed below.

(b) *Deposition on metals*

If a slightly roughened surface, such, for instance, as the normal surface of a coin, is sprayed, a very perfect negative is obtained, the smallest detail of the pattern being shown. On the other hand, if glass be sprayed, the surface of the metal film nearest the glass will not be smooth, but will contain a large number of regular pits. This fact has been noted by many workers, and Hoppelt gave as an explanation that the particles tended to glide on the smooth surface and lift before coming to rest. This explanation is difficult to follow and appears incorrect. The author believes that this phenomenon is due to the effect of surface tension. It is obvious that as the coin does not show this effect, the particle aggregates fill in every space in the first layer applied except, of course, the microscopical pores present in sprayed coatings. On the smooth surface of glass, possessing poor heat-conducting properties, there is nothing to impede the retraction due to surface tension. The first layer is covered before retraction has set in, then the first layer retracts, the second layer retracts a moment later, and so on, thus forming small pits.

It may be argued that it is unnecessary to bring in surface tension, because the same effect may be explained by ordinary contraction due to cooling. It will be seen from figure 14 that the area of the holes is too large to be covered by this explanation. The highest contraction allowed in making castings of most metals is about 2%. Taking area into account, this is insufficient to account for the phenomena described, although, of course, it is a contributing cause.

If the pistol be held at considerable distance from glass, the majority of the particles will be solid and plastic before reaching the glass. Splash effects will not be noted and the structure will appear to be built up of small, roughly circular units and the surface next to the glass will be nearly smooth. It should be noted, however, that adhesion is very small, and it takes a long time to form a coating, that is, such methods give a high percentage loss. It is impossible, however, to make a completely smooth coating in this way because some of the particles will be molten and act as described above.

§ 7. LOSS IN SPRAYING

This is a very difficult subject for investigation, because losses in spraying are dependent on so many factors, the following all having a very great influence on the actual loss figures :—

1. The movement of the pistol nozzle in relation to the surface sprayed.
2. The nozzle distance.
3. The type of surface.
4. The actual contours of the surface.

Nevertheless, it is possible to prove in every case that the loss with aluminium and its light alloys is very much less than with the other metals. Also, although tin has a very low melting point, the tendency is for lead to show the highest loss. In extreme cases, the loss with aluminium is one-tenth that with lead and about one-sixth of the loss with such metals as zinc and tin.

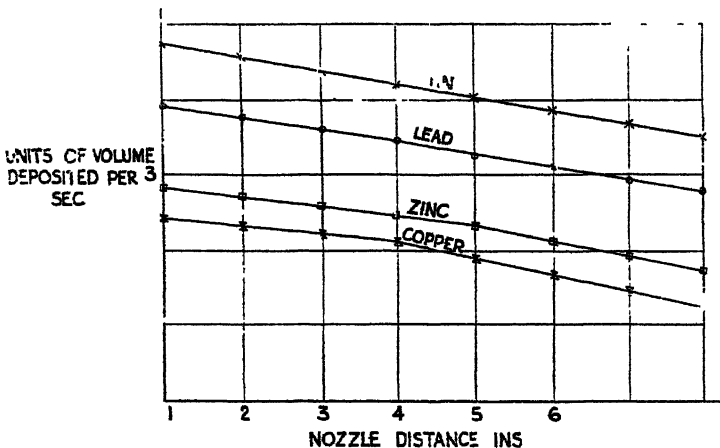


Figure 15. Curve showing volume of metal deposited at various nozzle distances

If one assumes that the particle size is not widely different, and it is assumed that the speed of the particles also is of the same order, it follows that the kinetic energy in each particle will be a function of its mass, and, therefore, the kinetic energy will bear some relationship to the reciprocal of the density. The aluminium particle would therefore appear to be likely to have one-quarter of the energy of the lead particle and one-third that of most other metals. In these circumstances any solid particle reaching the surface would be likely to bounce off much more easily in the case of lead than in the case of aluminium. This, therefore, would seem to indicate that the argument is sound, although many other factors will contribute to the final figures obtained for loss. For instance the proportion of total heat in an aluminium particle just solid would be four times that of a similar sized particle of lead. The metals of high melting point, such as steel and nickel, give a lower loss figure than would be indicated by considerations of kinetic energy alone, but the amount of heat in a solid particle nearing its melting point would be five times that of the lead particle. It would appear, therefore, that while it would be extremely difficult to express in a

mathematical formula the probable losses for any metal, the main contributing factors are the kinetic energy and the heat content.

If the volume of the metal deposited in unit time, under similar spraying conditions, be plotted against nozzle distance as the only variant, it will be found that for the metals of low melting point a straight line is obtained (figure 15), indicating that loss increases proportionately to increased nozzle distance over the normal range of spraying distance. In the case of the metals of higher melting point, this is true to a certain critical distance, after which a straight line of greater slope is obtained. In other words, beyond this distance loss increases more rapidly with higher nozzle distance. It is submitted that beyond this critical distance the majority of the particles are solid—hence the change in slope.

§ 8. DEPOSITS ON METAL AND SOME PRACTICAL OBSERVATIONS

We have considered the deposits made on smooth glass because the information gained leads to the assumption that surface tension has a considerable influence on their formation. Commercially, the deposits are usually applied to a roughened metal surface and the conditions are somewhat different. It is still true that there is no indication that the sprayed deposits actually wet the metal surface, but the conductivity of most metal surfaces will be considerably higher than the conductivity of a glass surface. Therefore, the particles will chill much faster than on glass, and the effects of surface tension will not be so apparent, as the metals will be solid before full retraction has taken place. If, therefore, surface tension is playing its part it would be expected that the adhesion of sprayed metal to an equally roughened surface of a metal of high conductivity would be somewhat less than that obtained on a metal surface of lower conductivity. This is, in fact, true, as it is more difficult to obtain high adhesion on cold copper or cold aluminium than it is on cold steel unless special arrangements are made with regard to surface roughness. Here again, however, the argument can be pressed too far because, if this is wholly true, the difficulty could be partly eradicated by heating the base metal, but this is not found to be particularly efficacious. It is obvious that the heating of the base may cause the formation of oxide films which may be torn apart by surface tension of the sprayed metal, and they in themselves will become loose and detachable from the surface.

The use of a suitably roughened surface such as that made by shot-blasting with a suitable grade of shot would seem to rely on the fact that the particles or agglomerates of particles, when they rest over the roughness of the surface, tend to retract in the way that has already been described, and this retraction will bind them on to each small mountain in the surface. As these mountains are more or less evenly spaced at very close intervals, retraction, instead of taking place over the whole sprayed surface, as in the case of smooth glass, is confined to shrinkage on to each raised portion, and, therefore, the effect of the shrinkage is to cause adherence to the surface as a whole rather than to tend to pull away in areas as is the case with glass. This effect, which is so important on shot-blasted surfaces, would not be found in the case of large contours such as a smooth screw-thread, because it will be seen that such large contours would be out of proportion to the size of the agglomerates of particles. In commercial practice,

however, the rough threading of shafts, for instance, does not leave a smooth surface within the thread, the surface itself being torn and so acting as a shot-blasted surface. Naturally, if the surface area is increased by such methods as screw threading, then the adhesion of the particles is increased per unit area of the original article, although it would not be increased in each very small area of the thread itself. As the thickness of the deposit increases, the shrinkage will go on in each layer, and obviously in the case of the grossly roughened surface, the shrinkage of the layers on top of each other will tend to bind it mechanically to the corrugation of the surface. In other words, on a normal shot-blasted surface the adhesion is largely due to surface-tension retraction effects, but on a screw surface this effect is increased by the retraction of the whole mass of metal on to the corrugations, so forming a mechanical lock.

One of the difficulties of the corrugated surface is that the solid particles passing down the margin of the sprayed cone do not so easily get carried away by the gases if they strike tangentially on the surface, and the result is that one may get porosities in the bottom of screw threads, for instance, due to the retention of the solid marginal particles. This is particularly obvious if the screw threads are too deep and too narrow.

If a pistol be held at an angle to the surface which is to be sprayed, there is a tendency to build up more quickly on the one side of the cone, and retraction, therefore, takes place preferentially towards the thickened section; this effect is likely to be cumulative. This is one of the causes of the rippling effect that is often shown when thick deposits are made under such conditions that the pistol cannot be applied at right angles to the surface to be sprayed.

It is known that the porosity of the coatings made by spraying increases with the nozzle distance (Fassbinder and Soularý, 1936; Ballard and Harris, 1936; Silfiant, 1937), and this is understandable in the consideration of the theory set out. Obviously, as nozzle distance increases, so will the number of solid particles reaching the surface, and the marginal effects will be more pronounced. At small nozzle distances, when the greater part of the deposit is formed by molten particles subject to chilling and retraction, the porosity is very small, and it would appear that the inclusion of solid particles (mainly due to marginal spray) is the major factor in the formation of the pores in sprayed metal.

Futhermore, at small nozzle distances, contraction will be greater, due to normal contraction of cooling metal and to the retraction effects which have been noted. If the coating is comparatively weak, the contraction effects will be enough to cause cracks as it cools. This effect is not often very apparent, but in the case of deposition of mild steel, if the speed of wire is increased too much by increasing the oxygen pressure of the flame, the coating becomes very oxidized and very friable, and if the metal is deposited on a shaft, the force of contraction becomes greater than the strength of the coating and cracks appear. It is suggested in some trade literature that to overcome this, nozzle distances must be increased to decrease contraction. It will be seen that this may not be a complete answer, as the coatings at greater nozzle distances quickly become more porous, and the structure will contain more solid particles. Under heavy loads this type or coating will tend to collapse and crack.

§ 9 CONCLUSION

This paper cannot claim to be exhaustive or to explain completely the phenomena connected with the art of metal spraying. Its purpose is to indicate some theoretical basis for observed facts, and to endeavour to describe some of the fundamentals involved. The author is deeply conscious of many problems yet unsolved. It is hoped that the ideas advanced may serve as a basis for further research.

ACKNOWLEDGMENTS

The author wishes to express his thanks to his laboratory staff, who have carried out the experimental work described, and to his colleagues on the Board of Metallisation, Ltd, for their encouragement of research.

REFERENCES

- ARNOLD, H., 1917. "Das Metallspritzverfahren: seine wissenschaftlichen, technischen und wirtschaftlichen Grundlagen." *Z. angew. Chem.* **30**, 209.
- BALLARD, W. E., 1943 *Brit. J. Phys. Med.* **6**, 21.
- BALLARD, W. E. and HARRIS, D. E. W., 1936 "Some observations on the metal spraying of copper" *Proceedings XII International Congress of Acetylene, Oxy-Acetylene and Allied Industries* (London), **5**, 1233.
- BURDEN, R. S., 1940. "Surface tension and the spreading of liquids", *Cambridge Physical Tracts*, p. 18.
- FASSBINDER, J. and SOULARY, P., 1936. "A contribution to the study of metal coating with the pistol". *Proceedings XII Internat Congr. Acetylene, Oxy-Acetylene Welding and Allied Industries* (London), **5**, 1219.
- HOTHERSALL, A. W., 1943. "Repair of worn or over-machined parts by electro-deposition" (*Symposium of papers on Reclamation of Worn Parts*). *Inst. Mech. Engrs.* November.
- INGHAM, H. and WILSON, K., 1944 "Procedures for testing metallising bond", *Iron Age*, pp. 44-49.
- MORCOM, R. K., 1914. "Metal spraying" *J. Inst. Met.* **12**, 116 124.
- ROLLASON, E. C., 1937. "Metal spraying-processes and some characteristics of the deposits." *J. Inst. Met.* no. 1, **60**, 35 66
- ROLLASON, E. C., 1939. *Metal Spraying* (Chas. Griffin and Co., Ltd.), p. 94.
- SILLIFANT, R. R., 1937. "Notes on some recent experiments in connection with the spraying of steel by the wire-fed metal-spraying pistol." *J. Iron Steel Inst.* **136**, 131.
- THORMANN, H., 1933. "Untersuchungen über das Metallspritzverfahren nach Schoop" (*Karlsruhe Badische technische Hochschule Fredericana*).
- TURNER, T. H. and BALLARD, W. E., 1924. "Metal spraying and sprayed metal." *J. Inst. Met.* no. 2, **32**, 291, 312
- WORTHINGTON, A. M., 1894. *Proc. Roy. Instn.* **14**, 289.

COMBINATIONS OF SPHERICAL LENSES TO REPLACE NON-SPHERICAL REFRACTING SURFACES IN OPTICAL SYSTEMS

By J. L. HOUGHTON,*

Research Laboratories, Kodak Ltd., Wealdstone, Harrow, Middlesex

Paper read to the Optical Group 18 September 1942 ; MS. received 3 October 1944

ABSTRACT. The substitution of lens systems of infinite focal length for non-spherical correcting plates is discussed, and formulae and examples are given for two- and three-component systems. It is pointed out that such systems may be used to introduce under- or over-correction of chromatic aberration and coma, in addition to spherical aberration, into a system. In the case of lens systems used in conjunction with spherical mirrors, the thicknesses and curvatures of the components are relatively small as compared with those found in more orthodox lenses of similar relative aperture, and the state of correction is good, though probably not so good as in the case where a non-spherical correcting plate is used.

§ 1. INTRODUCTION

DURING recent years several optical systems have been devised in which so-called *correcting plates* are used. The earliest and best-known example of such a system is the Schmidt camera, the basic principle of which has already been described many times. Although several ingenious methods have been developed for the manufacture of the plates, it is still a matter requiring skill, and consuming much time because the figuring of the plates is extremely heavy. In view of these difficulties, the manufacture of these plates in glass on a mass-production scale is not possible at the present time, and the alternative of using plastic materials does not appear to be very promising because the accuracy of the moulding is not yet sufficiently good. Spherical refracting systems on the other hand are easy to make in quantity and easy to test, so that the replacement of non-spherical by spherical surfaces may be advantageous.

The earliest example of such a substitution is to be found in the Mangin mirror, a divergent meniscus lens with a silvered convex surface, used in projection apparatus. If the power and shape of the lens are suitably chosen, the system may be freed from spherical aberration. The non-spherical system which is the counterpart of the Mangin mirror is the paraboloid reflector familiar to all users of astronomical reflecting telescopes. Apart from its use in projection apparatus, the Mangin mirror is not of great interest as the angular field it is capable of covering is very small.

Turning now to systems covering a large field, such as the Schmidt camera, the proposal is to replace the usual correcting plate by a lens system consisting of two or more lenses, and this lens system must possess the properties of a correcting plate. It must therefore have a focal length which is large in comparison

* Communication No. 988H from the Kodak Research Laboratories.

with the remainder of the optical system, it must be achromatic, and it must correct the spherical aberration of the remainder of the optical system. Other additional degrees of correction are possible in a lens system which are not obtainable in the conventional correcting plate. For example, the lens system may be over- or under-corrected chromatically.

§ 2 OUTLINE OF THE PRINCIPLES OF DESIGN

The approximate values of the powers and radii of curvatures of the components of the lens system are determined by the application of Seidel thin-lens equations. Most optical designers will prefer to use their own methods of design, but the following formulae from Chrétien (1938) have been found very convenient for the purpose.

Considering rays in the paraxial region, the incident height H_j of a ray on the j th lens will be proportional to the incident height H_1 of the same ray on the first lens in the system, and we may write

$$H_j = h_j H_1,$$

where h_j is the factor of proportionality. Similarly, for a principal ray, the incident height K_j of the ray on the j th lens is proportional to $\tan \theta_0$, where θ_0 is the angle of inclination of the ray to the optic axis and

$$K_j = -k_j \tan \theta_0.$$

If a component happens to be at the diaphragm, the coefficient k will clearly be zero. The quantities h_j and k_j may be calculated once the powers and separations of the components in the optical system are known, and will be used later in the calculation of the aberrations

The correcting lens system will be assumed to consist of two or more thin lenses in contact placed at the stop, and, if the system is achromatic, the following conditions are imposed:

$$\Sigma P = 0,$$

$$\Sigma \frac{P}{V} = 0,$$

where P represents the power of a component lens and V the dispersion figure for the glass of which it is made.

These equations are independent of the shapes of the components. The shape of the lens is expressed by

$$S = \frac{1}{2}(R_1 + R_2),$$

where R_1 and R_2 are the curvatures (reciprocal radii) of the two surfaces of the lens. If L_1 and L_2 are the object and image positions, real or virtual, on either side of a lens, a further variable T is defined as

$$T = \frac{1}{2} \left(\frac{1}{L_1} + \frac{1}{L_2} \right) = \frac{1}{2}(X_1 + X_2).$$

Quantities Q and E are then computed for each lens:

$$Q = AS^2 - 2BS + C$$

and

$$E = GS - H,$$

where, if n is the refractive index of the material composing the lens,

$$A = \frac{n+2}{n}P, \quad B = \frac{2n+1}{n}PT, \quad C = \frac{3n+2}{2}PT^2 + \left[\frac{n}{2(n-1)}\right]^2 P^3,$$

$$G = \frac{n+1}{n}P, \quad H = \frac{2n+1}{n}PT.$$

In the case of a mirror

$$P = 2R, \quad Q = \frac{1}{4}P^3 + XP^2 + X^2P, \quad E = -\frac{1}{2}P^2 - XP.$$

The spherical aberration s , coma c , and astigmatism a , of the complete system are then given by

$$s = \Sigma h^4 Q,$$

$$c = \Sigma [h^3 k Q - h^2 E],$$

$$a = \Sigma [h^2 k^2 Q - 2hkE + P].$$

The insertion anywhere in the system of a correcting plate or lens system of infinite focal length will clearly not affect the h , k , Q and E values for the other components of the system. If, in particular, such a plate or lens system is placed at the stop, the only aberrations of the complete system to be modified will be spherical aberration and coma, because, for the correcting lens system

$$\Sigma P = 0 \quad \text{and} \quad k = 0.$$

The guiding principle, therefore, in the application of these equations to optical systems of the type under review is to design the main optical system without the correcting lenses, so that the desired state of astigmatic correction is obtained, and to ignore the spherical aberration and coma. The insertion at the stop of suitable correcting lenses will then permit the correction of the outstanding spherical aberration and coma.

Take, as an example, the Schmidt system and consider the spherical mirror alone. Assuming a mirror of unit focal length imaging an infinite object :

$$P = 1, \quad h = 1, \quad X = 0, \quad Q = 0.25.$$

The astigmatism equation then becomes

$$0.25k^2 + k + 1 = 0,$$

which yields the solution $k = -2$.

The stop must therefore be placed two focal lengths in front of the mirror, that is, at its centre of curvature. If this value of k is substituted in the coma equation, the coma will also be found to be zero. Hence the correcting lens system is placed at the centre of curvature of the mirror, and is designed to have

$$Q = -0.25 \quad \text{and} \quad E = 0.$$

The simplest type of correcting lens system is one consisting of two lenses, which must clearly be of equal but opposite powers. The condition for achromatism indicates that the two lenses must be made of glasses having the same dispersion. Several such pairs are to be found in the optical glass catalogues; for example, Chance Brothers list a Soft Crown 515570, $N_D = 1.51508$, $V = 57.0$, and a Dense Barium Crown 621572, $N_D = 1.62081$, $V = 57.2$. Let the lenses be of powers P_1 and P_2 and let the front lens be of Soft Crown,

The spherical aberration contributed by the first component will be found to be

$$Q_1 = 2.32006 P_1 S_1^2 - 3.32006 P_1^2 S_1 + 3.24304 P_1^3$$

and, for the second component of Dense Barium Crown,

$$Q_2 = 2.23395 P_2 S_2^2 + 3.23396 P_2^2 S_2 + 2.76256 P_2^3.$$

Similarly, the coma contribution will be

$$E_1 = 1.66003 P_1 S_1 - 1.33002 P_1^2,$$

$$E_2 = 1.61698 P_2 S_2 + 1.30849 P_2^2,$$

the $h_3 k$ terms being zero since the lens system is at the stop.

Combining the spherical aberration and coma equations with the appropriate contributions of the mirror, which is of unit power,

$$S = 2.32006 P_1 S_1^2 - 3.32006 P_1^2 S_1 + 3.24304 P_1^3 - 2.23395 P_1 S_2^2 \\ + 3.23396 P_1^2 S_2 - 2.76256 P_1^3 + 0.25 = 0$$

$$C = 1.66003 P_1 S_1 - 1.33002 P_1^2 - 1.61698 P_1 S_2 + 1.30849 P_1^2 = 0.$$

There are three variables in these two equations, so that one can be fixed arbitrarily. A useful criterion to adopt is to make the front component symmetrical, that is $S_1 = 0$. Then from the coma equation

$$S_2 = -0.01331 P_1,$$

which is substituted in the spherical aberration equation to give a cubic equation for P_1 . The final solution is

$$P_1 = -0.83012, \quad S_2 = 0.01105.$$

The curvatures are then easily determined by the relations

$$R_1 = S_1 \pm \frac{P_1}{2(n_1 - 1)} \quad R_2 = S_2 \pm \frac{P_2}{2(n_2 - 1)},$$

the positive sign being taken for the front surface of the lens and the negative sign for the rear surface

The thin-lens specification is thus:

$$\begin{array}{ll} R_1 = -0.80582, & r_1 = -1.2410, \\ R_2 = 0.80582, & r_2 = 1.2410, \\ R_3 = 0.67963, & r_3 = 1.4714, \\ R_4 = -0.65753, & r_4 = -1.5208. \end{array}$$

The mirror, of course, will have a radius of -2.00 and be distant 2.00 behind the above lens system. When the requisite thicknesses and diameters are introduced, trigonometric ray-tracing will reveal the outstanding errors in spherical aberration and coma. A slight adjustment of the radius of the mirror and the distance of the mirror from the correcting lens will allow the solution to be corrected perfectly.

The two-lens system may also be used in a modification of the Schmidt, first proposed by Y. Vaisälä (1936). As is well known, the image field of the Schmidt is curved but anastigmatic, the radius of curvature being equal to the

focal length. By the introduction of some astigmatism the tangential field may be flattened in the manner recommended by Conrady (1929). It may be shown that the amount of astigmatism required for this purpose is one-third of the Petzval sum of the system. The Petzval sum of the correcting lens system is small compared with that of the mirror, and may be neglected. The Petzval sum of the mirror is equal to its power. Hence the equation for the stop position to give a flat tangential field is

$$\frac{1}{3} = \frac{k^2}{4} + k + 1,$$

the solution of which is $k = -3.1547$ or -0.8453

The first solution is not attractive, as a large mirror would be required to avoid vignetting, and the second requires a diagonal reflector or a central hole cut in the lens, because the correcting lens will then be between the mirror and image plane. If, however, $k = -1$ is taken, that is, if the stop is at a distance equal to the focal length of the mirror, the astigmatism is 0.25, so that the tangential field is slightly "round". The coma is not zero for a stop in the focal plane; hence, if a correcting plate is used, it is necessary to figure the mirror to remove the coma. If a lens system is substituted for the usual correcting plate, the coma may be corrected in the lens system instead of by figuring the mirror. Details of the calculation of such a lens combination will not be given, as it follows closely the example already given.

The application of correcting lens systems is not of course limited to reflecting optical instruments, but may be introduced into refracting systems. As an example, consider a thin singlet lens. By choosing the shape of the lens and the position of the stop, such a system may be rendered flat-fielded and free from coma, but the spherical aberration cannot then be corrected. This correction may be carried out by inserting a correcting lens at the stop which is designed to be free from coma.

The above example is of academic interest only, when regarded from the point of view of lens designing; three components have been used to produce a system which would probably be inferior in performance to an anastigmat triplet of the Cooke type. The principle involved, however, is applicable to lens systems of various and more complicated types.

In the case of a Schmidt camera, a further lens system, suggesting itself as a substitute for the orthodox correcting plate, consists of three components, a pair of equiconvex and similar lenses with an equiconcave lens, the power of which is equal and opposite to the sum of the powers of the convex lenses, between them. A considerably wider choice of glass is possible with such a triplet formation than is the case with the doublet. Any pair of glasses may be used, including the same glass throughout, the only condition being that the glasses are of identical dispersions. The symmetrical system is free from coma regardless of the powers of the components, which can easily be verified from the formulae already given by the substitution of

$$S_1 = -S_3, \quad S_2 = 0 \quad \text{and} \quad T_1 = \frac{P_1}{2}, \quad T_2 = 0, \quad T_3 = -\frac{P_1}{2}.$$

Trial solutions will show that there is not much advantage to be gained by departing from the symmetrical shapes of the components and that a high refractive index

is advantageous in reducing zonal spherical aberration. If all the lenses are made of the same glass, the formulae become particularly simple

Assuming $S_1 = S_2 = S_3 = 0$; $P_1 = P_3$; $P_2 = -2P_1$, the equation for P_1 is

$$\left[\frac{3n+2}{2n} - \frac{3}{2} \frac{n^2}{(n-1)^2} \right] P_1^3 + 0.25 = 0.$$

For $N=1.8$, $P_1=0.3561$ and $P_2=-0.7122$.

For a mirror of unit focal length the thin-lens specification for the lens system will be found to be

$$\begin{array}{ll} r_1 = 4.4932, & r_4 = 2.2466, \\ r_2 = -4.4932, & r_5 = 4.4932, \\ r_3 = -2.2466, & r_6 = -4.4932. \end{array}$$

Comparing these values for the radii with those obtained for the two-lens system, it will be seen that the curvatures are much smaller. This allows a larger numerical aperture to be used, and aperture ratios of about $F/0.6$ are attainable with the triplet formation as compared with about $F/1.4$ for the doublet, while still retaining adequate image sharpness.

The triplet systems are not necessarily limited to coma-free formations; for instance, by a suitable redistribution of the powers of the components with possibly differing shapes, coma may be introduced to correct that in some other part of the complete system. The design of such systems will, of course, be a little more complicated than is the case with the symmetrical systems, but it will again follow the formulae outlined earlier in this paper.

In the numerical examples given, the correcting lens systems are over-corrected for spherical aberration and are achromatic, but there is no reason why spherical under-correction may not be obtained if it is required. Chromatic over-correction may be desirable in certain cases; for example, an over-corrected system may be used in a Schmidt type of instrument where the field is flattened by an oil-immersion aplanatic lens similar to the front lens of an oil-immersion microscope objective. The system would consist of a concave lens of high-dispersion glass between two convex lenses of low-dispersion glass. The oil-immersion lens would then fulfil the double function of flattening the field and increasing the relative aperture.

Owing to the shallow curves and thin components, lens-mirror combinations of the types described in this paper are faster and more highly corrected than the more conventional lens systems, but they are probably slightly inferior in these respects to those systems using a correcting plate, since all systems using spherical surfaces are afflicted more or less with zonal aberrations. A few trigonometrically corrected examples will be found in B.P. 546,307, including the doublet and triplet Schmidt types and the flat-fielded type briefly described in this paper.

REFERENCES

- CHRÉTIEN, H., 1938. *Cours de Calcul des Combinaisons Optiques* (Revue d'Optique, théorique et instrumentale, Paris), pp. 228, 236, 254, 349, 477, 480, 616, 790.
 CONRADY, A. E., 1929. *Applied Optics and Optical Design* (Oxford University Press), pp. 291-95.
 VAISALA, Y., 1936. *Astr. Nachr.* 259, 198-203.

The following references, for which I wish to thank Dr. Kingslake, were brought to my notice after the above paper had been read to the Optical Group. They are offered as a small bibliographical supplement:—

Zeiss describe in B.P. 482,874 a mirror-lens system in which the reflecting element is a meniscus divergent lens silvered on the convex surface. In this case, the mirror-lens contributes strongly to the spherical correction of the complete system; in fact, it is practically a Mangin mirror. Gabor (B.P. 544,694) replaces the correcting plate of the conventional Schmidt camera by a meniscus divergent lens of which an annular ring only is used, and claims that effective apertures less than $F/1.0$ can be obtained and fields of about 23° square can be covered. Martin, Flugge and Roll (U.S.P. 2,229,302) describe systems for use with curved image surfaces such as would be found in fluorescent screens of cathode-ray tubes in television apparatus. Cassegrain systems are described by Acht (U.S.P. 1,967,214 and U.S.P. 1,967,215) and by Tate (B.P. 426,539). A very important paper on the replacement of non-spherical surfaces by spherical surfaces has recently been published by D. Maksutov, "New Catadioptric Meniscus Systems", *J. Opt. Soc. Amer.* **34**, 270-284 (May 1944). Maksutov points out that a monocentric meniscus lens may be used instead of a corrector plate, and gives several examples of systems using such a lens, together with a detailed analysis of the aberrations. Empirical formulae for the design of systems of the Schmidt type are given.

THE NON-REFLECTING TERMINATION OF A TRANSMISSION LINE

By D. H. SMITH,
Woolwich Polytechnic

MS. received 13 October 1944

ABSTRACT. In a recent paper, Willis Jackson and Huxley have presented some experimental results which appear not to agree with the theory of the non-reflecting termination of a transmission line when the latter is terminated by a thin film. The present paper shows that their results may be explained by the partial transparency of a very thin conducting film to electromagnetic waves, and that there is no real disagreement with the usual theory

§1 INTRODUCTION

THE simple theory of a transmission line shows that if it is terminated by an impedance which is equal to the characteristic impedance of the line, waves arriving at the termination will not be reflected and the line will behave as if it were infinitely long. If the losses may be neglected, as they may be in many practical applications of transmission lines at radio-frequencies, the characteristic impedance reduces to a pure resistance. It is sometimes necessary in laboratory work to provide the line with a non-reflecting load, and it becomes of interest to consider the physical nature of the conductor which is

needed. In the case of a line whose members are two similar parallel wires, it is apparently easy to place a resistance of the required value across the end of the line, in the form of a strip of suitable material, but when a coaxial line is used, this would appear to lead to an undesirable lack of symmetry. The conductor which naturally suggests itself is a disc or film of conducting material extending from the inner member to the outer member of the line, and it is easy to show that the resistance of such a film is

$$R = \frac{1}{2\pi gh} \ln \frac{r_2}{r_1} \text{ ohms,} \quad \dots\dots(1)$$

where g is the conductivity of the material, h is its thickness, and r_1 and r_2 are the radii of the inner and outer conductors respectively. The characteristic impedance of a loss-free coaxial line with air-spacing is

$$Z_{00} = 60 \ln \frac{r_2}{r_1} \text{ ohms,} \quad \dots\dots(2)$$

and R will therefore be equal to Z_{00} if

$$1/gh = 2\pi \times 60 = 377 \text{ ohms.} \quad \dots\dots(3)$$

The quantity $1/gh$ is the resistance between opposite edges of a square piece of the film whose side is 1 cm., and in the current literature it is sometimes called, not too happily, the surface resistivity. The quantity 377 ohms is also the value in free space of $377\sqrt{\mu/k}$, where μ is the permeability and k is the dielectric constant of the medium between the conductors; it has been called by Schelkunoff (1938) the *intrinsic impedance of free space*. It is usually assumed that, in these terms, there will be no reflection if a line is terminated with a thin film whose surface resistivity is equal to the intrinsic impedance of free space, but in a recent paper Willis Jackson and Huxley (1944) have described experiments which show that this is not the case. These authors found that when a coaxial line whose characteristic impedance was 76 ohms was terminated by a graphite film deposited on bakelized paper, of resistance 77 ohms between the inner and outer conductors, a pronounced system of standing waves was formed, and an analysis of the standing-wave pattern showed that the film behaved as a complex impedance $23.1 - j20.1$ ohms, equivalent to a resistance of 40.5 ohms in parallel with a capacitative reactance of 46.5 ohms.

It was found, however, that the line could be correctly terminated by continuing it beyond the film to a distance of $0.88 \lambda/4$, where λ is the wave-length, and closing the extension by a short-circuiting disc. As the authors point out, there appears to be widespread misunderstanding of this problem. It has also been discussed by Howe (1944) in an editorial article in the *Wireless Engineer*.

§ 2 THE REFLECTION OF A PLANE ELECTROMAGNETIC WAVE BY A THIN CONDUCTING FILM

Neither in the original paper, nor in the discussions which have followed it, does attention appear to have been directed to the significant fact that a film so thin is markedly transparent to electromagnetic waves falling upon it, as was shown by Heaviside (1892).

There will be no loss of generality if we consider a transmission line made up of two plane-parallel slabs of perfectly conducting material, a cm. apart and b cm. wide, flanked by guard-plates separated from them by infinitesimal gaps (figure 1). The electric and magnetic lines of force are straight lines, and the geometrical relationships are peculiarly simple. The capacity per cm. of such a line is $C = \kappa \cdot \frac{b}{a}$ farads, where $\kappa = \frac{k}{36\pi \times 10^{11}}$, κ being the dielectric constant of the medium.

The inductance per cm. is $L = \bar{\mu} \cdot \frac{a}{b}$ henries, where $\bar{\mu} = 4\pi\mu \times 10^{-9}$.

The characteristic impedance is therefore,

$$Z_{00} = \sqrt{L/C} = \sqrt{\left(\frac{\bar{\mu}}{\kappa}\right)} \times \frac{b}{a} = 377 \frac{b}{a} \text{ ohms.} \quad \dots (4)$$

Let the line be semi-infinite, and let a film of any "surface resistivity" R_0 be placed across it at any point, with extensions across the guard-plates on each side.

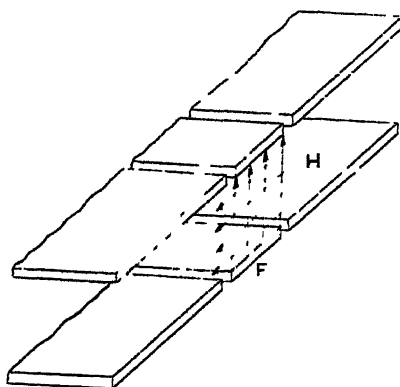


Figure 1

The film must necessarily be very thin. Thus with graphite, which has been used in practice, g is of the order $1250 \text{ ohm}^{-1} \text{ cm.}$, and if $R = 377 \text{ ohms}$, h will be of the order $2 \times 10^{-6} \text{ cm.}$ With better conductors, h must be correspondingly smaller.

Consider now a plane electromagnetic wave, guided by the transmission line, which meets the film, and let H_1 amperes per cm. and F_1 volts per cm. be the strengths of the magnetic and electric fields in the wave.

In terms of these units, Maxwell's equations may be written

$$\oint H \cos \theta \, ds = \iint \left(\kappa \frac{dF}{dt} + I \right) dS, \quad \dots (5)$$

$$\oint F \cos \theta \, ds = \iint \bar{\mu} \frac{dH}{dt} dS, \quad \dots (6)$$

where in (5) $H \cos \theta$ is the component of H tangential to the element ds of a closed curve and dS is an element of any area bounded by the curve. $\kappa \frac{dF}{dt}$ is the displacement current density and I is the conduction-current density.

A similar terminology applies to (6).

In the plane waves considered, the general relationships

$$F = \bar{\mu} v H, \quad \dots\dots(7)$$

$$H = \kappa v F \quad \dots\dots(8)$$

exist between the fields, where $v = \sqrt{1/\bar{\mu}\kappa}$ is the velocity

At the first face of the film, the tangential component of the electric field must be continuous. Let (F_3, H_3) be the fields in the reflected wave, and let (F_2', H_2') be the fields in the wave transmitted into the film (figure 2).

Then from the equation of continuity,

$$F_1 - F_3 = F_2' \quad \dots\dots(9)$$

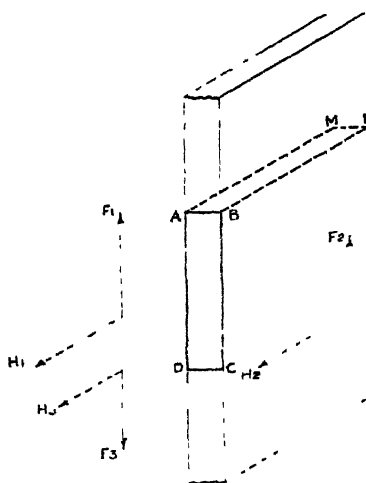


Figure 2.

Let (F_2, H_2) be the fields at the second face of the film. Applying equation (6) to the rectangle ABCD, we obtain

$$F_2 + F_3 - F_1 = \bar{\mu}_1 \int_0^h \frac{dH}{dt} dh = \bar{\mu}_1 \kappa_1 v_1 \int_0^h \frac{dF}{dt} dh = \frac{1}{v_1} \int_0^h \frac{dF}{dt} dh, \quad \dots\dots(10)$$

where $\bar{\mu}_1$, κ_1 , and v_1 relate to the material of the film, and (F, H) are the fields at any point in the film.

If we consider sinusoidal waves, the right-hand side of (10) is of the order $\omega F h / v_1$, and if the frequency is of the order 10^8 c./s., and h is of the order 10^{-6} cm., the order of this quantity is 2×10^{-8} F.

Therefore F_2 only differs from $F_1 - F_3$ by an extremely small quantity, and, without serious error, we may say that $F_2 = F_2'$.

$$\therefore F_1 - F_3 = F_2. \quad \dots\dots(11)$$

Thus F is not attenuated. It is otherwise with H . Equation (5) may now be applied to the rectangle ABLM, and similar reasoning shows that the contribution

of the displacement current to the right-hand side is negligible. The conductive-current term is, however, important, and is given by

$$\iint I dS = F_2 \cdot gh \cdot \text{AM.}$$

Then $(H_1 + H_3 - H_2) \text{AM} = F_2 gh \cdot \text{AM.}$

$$\therefore H_1 + H_3 - H_2 = F_2 gh. \quad \dots\dots(12)$$

From (11), (12), (7) and (8) we obtain

$$\frac{F_3}{F_1} = \frac{gh}{gh + 2\kappa v}, \quad \dots\dots(13)$$

$$\frac{F_2}{F_1} = \frac{2\kappa v}{gh + 2\kappa v}. \quad \dots\dots(14)$$

Now $\kappa v = \frac{1}{\mu v} = \frac{1}{Z_{00}} \times \frac{b}{a}$ and $gh = R_0$, the surface resistivity of the film. In

terms of Z_{00} and R , the total resistance of the film,

$$\frac{F_3}{F_1} = 1 / \left(1 + \frac{2R}{Z_{00}} \right), \quad \dots\dots(15)$$

$$\frac{F_2}{F_1} = 1 / \left(1 + \frac{Z_{00}}{2R} \right). \quad \dots\dots(16)$$

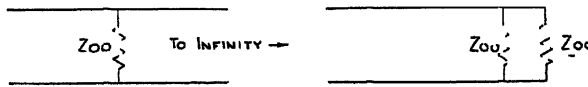


Figure 3

If R is made equal to the characteristic impedance of the line, we shall have

$$\frac{F_3}{F_1} = \frac{1}{3}, \quad \frac{F_2}{F_1} = \frac{2}{3},$$

and the amplitude of the transmitted wave is thus two-thirds that of the incident wave. A somewhat more complicated analysis shows that for a coaxial line or a twin line, the corresponding voltages are in the ratios given by equations (15) and (16).

The above calculation applies only to the case where the line extends beyond the film to infinity, so that the transmitted wave can be guided on. In view of the mistrust shown by Willis Jackson and Huxley of the use of the impedance concept, it may be noted that the result obtained is precisely that given by a calculation in terms of impedances. For if a resistance Z_{00} ohms be connected at any point A across a semi-infinite line (figure 3), the impedance presented to a wave arriving at A is made up of Z_{00} in parallel with the input impedance Z_{00} of the rest of the line. It is therefore $Z_i = \frac{Z_{00}}{2}$, and the reflection coefficient is $\frac{Z_i - Z_{00}}{Z_i + Z_{00}} = \frac{1}{3}$.

In the experiments of Jackson and Huxley, the film was placed across the end of the coaxial line, and in such a case it is clear that a wave will be transmitted

through the film and radiated away. The terminating impedance will then be a parallel combination of the resistance of the film and the radiation impedance of the annular aperture, a quantity not easily predictable.

§3 THE NON-REFLECTING TERMINATION

It is of interest to examine, from the aspect of wave-propagation, the termination which Jackson and Huxley found to be effective in suppressing reflection. This was a short-circuited extension of the line, beyond the film.

Let x be the length of the extension. As before, let F_1 be the amplitude in the wave advancing towards the film. Consider any film for which pF_1 is the amplitude of the transmitted wave, and $(1-p)F_1$ is the amplitude of the reflected wave where $0 < p < 1$. The transmitted wave is perfectly reflected by the short-circuiting disc, with reversal of the sense of the electric field, and this reflected wave is partly transmitted and partly reflected when it returns to the film. Successive reflections take place in this way at the boundaries of the extension piece, and if we take the phase of the first reflected component as standard phase, the

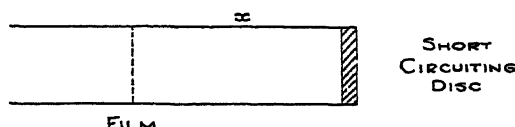


Figure 4.

electric field in the disturbance travelling back from the film towards the origin is the sum of the infinite series

$$F = F_1 \{ (1-p) \exp. j\omega t + p^2 \exp. j(\omega t - 2qx) + p^2(1-p) \exp. j(\omega t - 4qx) + p^2(1-p)^2 \exp. j(\omega t - 6qx) + p^2(1-p)^3 \exp. j(\omega t - 8qx) + \dots \},$$

where $q = 2\pi/\lambda$.

$$F = F_1 \exp. j\omega t \frac{(1-p) - (1-2p) \exp. (-2jqx)}{1 - (1-p) \exp. (-2jqx)}.$$

The square of the amplitude (F_0) may be obtained by multiplying by the conjugate complex quantity. It is

$$F_0^2 = \frac{(1-p)^2 + (1-2p)^2 - (1-p)(1-2p)^2 \cos 2qx}{1 + (1-p)^2 - (1-p)^2 \cos 2qx}.$$

The condition for this to vanish is

$$\cos 2qx = - \frac{(1-p)^2 + (1-2p)^2}{(1-p)(1-2p)},$$

If $p = \frac{2}{3}$, which is the value obtaining when the resistance of the film is equal to the characteristic impedance of the line, as in Jackson and Huxley's experiments, this reduces to $\cos 2qx = -1$.

Therefore $x = (2P+1) \frac{\lambda}{4}$, where $P = 0, 1, 2 \dots$,

Thus a suitable film, backed by a short-circuited extension of the line, one-quarter wave-length long, will correctly terminate the line, and the energy of the incident wave will be completely dissipated in the course of the successive passages through the film. The same result could be obtained by terminating the line with the film alone and using matching stubs, but the asymmetry introduced by stubs in the case of a coaxial line would be undesirable, and Jackson and Huxley's device is much to be preferred. The departure of the length of their extension ($0.88 \lambda/4$) from the theoretical value of $\lambda/4$ may be due to the sheet of bakelized paper on which the film was deposited, and to the method of mounting.

The result is again in agreement with that given by the concept of impedance. The input impedance of the short-circuited extension of length $\lambda/4$ is infinite, and the impedance presented by the film and the extension to the oncoming wave is therefore Z_{00} . This agreement arises from the fact that the currents and the voltages, in terms of which the impedances are defined, are themselves variables derived by integration from the intensities of the magnetic and electric fields respectively; the latter are the fundamental entities. As correct relationships, satisfying Maxwell's equations, exist between the fields, so must they between the impedances. The use of impedance, however, obscures the physical realities. The subject is one which lends point to the aphorism of Hansen (quoted by Slater, 1942]: "The idea of impedance cannot be used as a substitute for thought".

Jackson and Huxley's terminating device is non-reflecting for one frequency only. The ideal non-reflecting termination for a coaxial line would be a length of distortionless line, with high attenuation, having the same physical dimensions and characteristic impedance as the line to be terminated, but it is doubtful whether this could be set up with known materials.

REFERENCES

- HEAVISIDE, 1892. *Electrical Papers*, vol. II, p. 385.
 HOWE, 1944. *Wireless Engr.* 21, 409.
 JACKSON and HUXLEY, 1944. *J. Instn. Elect. Engrs.* 91, 105.
 SCHELKUNOFF, 1938. *Bell Syst. Tech. J.* 17, 24.
 SLATER, 1942. *Microwave transmission*, p. 79.

A STUDY OF THE COMPARATIVE METHOD OF DETERMINING GASEOUS REFRACTIVITIES*

By E. C. CRAVEN

MS received 25 October 1944 ; in revised form 14 December 1944

ABSTRACT. It is shown that the comparative method for the determination of refractive indices by white-light interferometry yields the ratio of the group refractivities and not the ratio of the phase refractivities.

The effective wave-length of the ordinary gas-filled lamp was found to be close to that of sodium light. For ordinary gases, however, any wave-length in the same region may be taken without serious error.

Observation of the shift of the central achromatic fringe due to differences of dispersion in the two paths enables the difference of dispersion to be measured. This enables the group index to be corrected and the phase index calculated at any point in the visual range of the spectrum.

The method has been applied to several common gases and to mixtures of the vapours of volatile organic solvents with air, and appears to give results comparing not unfavourably in accuracy with those obtained by the absolute method in monochromatic light.

§ 1. INTRODUCTION

THE interferometric determination of the refractive index of a gas or vapour may be carried out either by an absolute method or by a comparative method. In the former, monochromatic light is used and the band-shifts for various changes of pressures are observed. By extrapolation, the band-shift for the pressure change from zero to one atmosphere is calculated, and hence the refractive index. The appropriate corrections are applied to give the result at N.T.P.

In the comparative method, instead of counting the fringe difference, optical balance is restored in effect by introducing a suitable thickness of a comparison substance, usually air or glass. The variation of path in the comparison gas is most easily effected by varying its pressure in a tube of constant length. This method of varying the gas pressure has the advantage of giving a substantially automatic correction for normal changes of room temperature and pressure. When a glass compensator is used, both temperature and pressure of the gas have to be taken into account.

Now the measurement of the apparent relative retardation of two gases merely involves the determination of the ratio of the two gas pressures at optical balance, a determination which can obviously be made with considerable precision. The comparative method is more rapid and convenient than the absolute method. For the determination of the concentration of a solvent vapour in air, for example, there is no method which has anything like the combined speed and precision of the interferometric one.

* Abstracted from a thesis approved by the University of London for the Ph.D. degree.

There is, however, no way of identifying the central fringe of the interference system in monochromatic light, so that white light has to be employed to give an achromatic central fringe, and thus to enable restoration of optical balance to be observed. It appeared desirable to examine the possibilities of the comparative method, and to find what differences are introduced by the necessary substitution of white light for monochromatic light.

As a comparison gas there can be no doubt that dry CO_2 -free air is the most convenient to employ, as has already been suggested by Werner (1925) even for mixtures containing no air. Where mixtures of air and vapour are concerned, the use of air commends itself even more strongly.

The degree of accuracy aimed at in the present work was of the order of about 1 part in 1000 in measurements of $(n-1)$, which should be attained by pressure measurements on the air compensation tube within a few tenths of a millimetre of mercury.

The chief source of uncertainty in the comparative method arises from the necessary use of white light for the identification of the central fringe. A theory is developed below dealing with the corrections introduced by the shift of the achromatic fringe due to dispersion. It was proposed to verify this theory by measurements on pure gases, the optical constants of which had been reasonably well established by previous workers, taking into account any corrections arising from departure from the ideal gas laws, and finally to apply the method to mixtures of solvent vapours with air. The published values for solvent vapours are in general rather old.*

The problem of the determination of refractive index by white-light interferometry has been dealt with to some extent by Adams (1915) and by Clack (1925), both of whom were concerned with aqueous solutions, and used tilting-plate compensators. The present treatment applies primarily to the (gas) pressure-variation method of compensation, but can readily be extended to cover the tilting plate or any other method.

Apart from an observational refinement due to Clack, the present treatment is in principle exactly the same as the theory of the method of the determination of the index of refraction and dispersion of a glass plate as given by Mann (1902).

§2 THE DENSITY CHANGE FOR BALANCE

The density change required to obtain optical balance in non-homogeneous light may be calculated as follows :—

Suppose two tubes of unit length, containing two gases of refractive indices n_1 and n_2 and densities ρ_1 and ρ_2 . Optical balance is obtained by increasing the pressure of the second gas until its density is $f\rho_2$, where f is a density ratio. Applying the Gladstone-Dale law, the change rate of phase or order with wave-length will be zero at the centre of the system, when

$$f = (n_1 - \lambda \frac{dn_1}{d\lambda} - 1) / (n_2 - \lambda \frac{dn_2}{d\lambda} - 1).$$

It can readily be shown by assuming some dispersion formula, such as

$$n - 1 = A + B\lambda^{-2},$$

* e.g. the 1941 edition of Kaye and Laby's *Tables* quotes values mostly determined about 1878,

or by graphical methods that the ratio

$$(n - \lambda \frac{dn}{d\lambda} - 1)/(n - 1) = 1 - (\lambda \frac{dn}{d\lambda})/(n - 1)$$

is very closely equal to $1 + 3\omega$, where ω is the dispersive power expressed as

$$\omega = (n_F - n_C)/(n_D - 1).$$

§ 3 VERIFICATION OF GROUP-INDEX FORMULA

The ratio f may be obtained in practice from the ratio of the initial and final pressures by making the appropriate correction, determined if necessary by the method of limiting densities.

The replacement of n in monochromatic light by $n - \lambda \frac{dn}{d\lambda}$ in non-homogeneous light was pointed out long ago by Rayleigh, Cornu and others. As emphasized, especially by Schuster for the analogous case of a plate of some dispersive solid, the same result is obtained by considering that what is involved is the *group velocity* of the trains of waves resulting from white-light impulses. For the velocity $V = c/n$ of monochromatic light in the medium must be substituted the smaller group velocity $U = V - \lambda \frac{dV}{d\lambda}$. The path retardation in each tube of gas at the original density is given by the numerator and denominator in the expression for f above.

By analogy with the ordinary phase refractive index $n = c/V$, the quantity $c/U = n - \lambda \frac{dn}{d\lambda}$ may be termed a "group index". A summary of applications of this quantity is given later, and the symbol \bar{n} is tentatively proposed for it.

Although the above theory is old, the author is not aware of any measurements which have been made to confirm it with the exception of the classical determination of Michelson in 1883 of the velocity of white light in water and CS_2 by the rotating-mirror method, and further direct determinations of velocity in liquids more recently by others (Gutton, 1911; Houstoun, 1944).

The following measurements may, therefore, be of special interest. In a Rayleigh refractometer with 50-cm. tubes (Craven, 1939) the change in air pressure necessary to balance equal films (about 0.02 cm.) of various liquids in turn was determined. The cell was made of ordinary plate glass and was mounted

Fluid	p	n_D	\bar{n}	$\frac{\bar{n}-1}{n-1}$	$1+3\omega$
Air	0	1.000	1.000	—	—
Water	353	1.333	(1.351)	(1.054)	(1.054)
Methyl alcohol	348.5	1.329	1.346	1.052	1.050
Ethyl alcohol	381	1.362	1.379	1.047	1.050
Acetone	383	1.359	1.380	1.059	1.059
Benzene	557	1.501	1.554	1.105	1.100
Anilin	662	1.586	1.658	1.123	1.127
Carbon disulphide	731	1.628	1.727	1.158	1.162

p = the air pressure in mm of Hg required to restore balance.

n_D = refractive index for sodium light.

\bar{n} = apparent index for white light, i.e. the group index

The observed ratio of $\bar{n} - 1$ to $n - 1$ is in reasonable agreement with the value of $(1 - 3\omega)$, calculated from published values of n_F , n_D , and n_C , and given in the last column

in front of the gas tubes, and the fiduciary air system* was formed, as usual, above the tubes.

The temperature was 20° c. throughout, and the results are calculated by comparison with water as the standard substance.

No doubt with a cell of optically worked parts and attention to a further small correction mentioned in a later section, even closer results would be obtained.

The number \bar{n} represents the reciprocal of the velocity of white light in the medium as compared with the velocity of light in free space. Michelson's direct determinations in 1883 were made with a tube 3 metres long, whereas in the present work the film thickness was barely 0.2 mm. The result obtained by Michelson for CS_2 was 1.758, and the temperature appears to have been 0° c.

Further experiments were made in which 1 cm. of water was optically balanced by a suitable thickness of glass, and others in which microscope cover-glass was balanced by air pressure. The results were all closely in agreement with the group-index principle. Measurements were also made on flakes of mica and the disappearance of the fringes with $\lambda/2$ plate and a $3\lambda/2$ plate was noted.

§ 4. SOME GENERAL THEORETICAL POINTS

The present work appears to confirm that, to a first approximation, interferometry in white light yields the group index and not the ordinary phase index. The question has been asked by Williams (1932) at what breadth of source we pass from phase index to group index. The answer appears to be that any real source will give group-index results by alignment of the achromatic centres. The position of the achromatic centre, however, quickly becomes obscure with reduction of spectral width of the source.

Some general applications of the group-index idea may be of interest. The well-known Rayleigh formula for group velocity is based on two contiguous wave-lengths, λ and $\lambda + d\lambda$. With a source of greater width, λ becomes the "dominant" or effective wave-length and the group index $n - \lambda dn/d\lambda$ becomes an average value, and it is therefore tentatively proposed to denote it by \bar{n} .

According to the group-index idea, a thin pulse of non-homogeneous light passing through any real medium is spread into a train of waves and the condition for achromatism in any system is that on every such train there shall be superposed an equal and opposite train. By such considerations many of the ordinary formulae of optics may be derived and expressed in terms of the group index \bar{n} and the ordinary phase index n :

Retardation of slab	$t(\bar{n} - 1)$
Interferometric balance	$t_1(\bar{n}_1 - 1) = t_2(\bar{n}_2 - 1)$
Resolving power of echelon	$Nt(\bar{n} - 1)\lambda$
Resolving power of prism	$t(\bar{n} - n)\lambda$
Condition for thin achromatic prism	$t_1(\bar{n}_1 - n_1) = t_2(\bar{n}_2 - n_2)$
Condition for thin achromatic doublet	same as prism.

In each case t is the effective thickness, i.e. difference of thickest and thinnest parts in use, and N is the number of elements in the echelon.

* The use of the fiduciary system is often ascribed to Haber and Lowe, but is in fact explicitly described by Lord Rayleigh (*Sci. Papers*, 4, 221).

It may be asked why the group index appears so little in ordinary optical measurements. It is suggested tentatively that this is because in general such measurements are concerned with angular deviation. As a dimensionless quantity, an angle has no relation to the mode of propagation of light in the medium, neither does it give any direct information as to the velocity of propagation. The group mechanism relates only to effects in the line of sight, and these are seldom examined.

§ 5. DISPLACEMENT OF THE ACHROMATIC FRINGE

In the expression for the change in density ratio required to obtain optical balance in white light viz.,

$$f = (\bar{n}_1 - 1)/(\bar{n}_2 - 1),$$

if a simple Cauchy dispersion formula $(n - 1) = A + B\lambda^{-2}$ be accepted for each gas, then the corresponding group refractivity will be

$$\bar{n} - 1 = A + 3B\lambda^{-2}.$$

Applying this, for example, to the case of the relative refractivities hydrogen/air it will be found that

Relative refractivity in sodium light = $1394/2926 = 0.4763$.

Relative refractivity in white light = $1453.3016 = 0.4817$.

The above difference in relative refractivity is equivalent to $0.0054 \times 462 = 2.5$ fringes of Na light in a 100-cm. tube at 20° C. In monochromatic light there would be an equal number of fringes in each of the paths through the tubes at optical balance, whereas in white light, owing to the dispersion of the medium, there are unequal numbers. This is manifested by a wandering of the central achromatic band from the position it had under the original symmetrical conditions.

§ 6. THE DISCONTINUITY RESULTING FROM THE SHIFT OF THE ACHROMATIC FRINGE

In the above calculation it is assumed that in restoring optical balance, the optical centres of the fiduciary and gas-tube fringe systems are brought to coincidence, whereas in actual fact what is done is to align the achromatic band of the reference system with the most nearly achromatic band of the gas system. It has already been shown by Clack (*loc. cit.*) that the effect of this natural technique is that at balance the numbers of periods in the two paths are either equal or differ by an integral number. If, therefore, starting with the two gas tubes full of the same gas A, a second gas B of higher refraction and dispersion is gradually substituted in one tube, the resulting curve is as indicated diagrammatically by the full lines in the figure.

The line OXF' gives the balance in monochromatic light corresponding to the equation

$$f' = (n_B - 1)/(n_A - 1).$$

In white light there is a discontinuity at X, the achromatic centre having shifted half a fringe, so that there are two equally achromatic fringes in the gas system. This condition is frequently observed in practice. At Y the displacement is

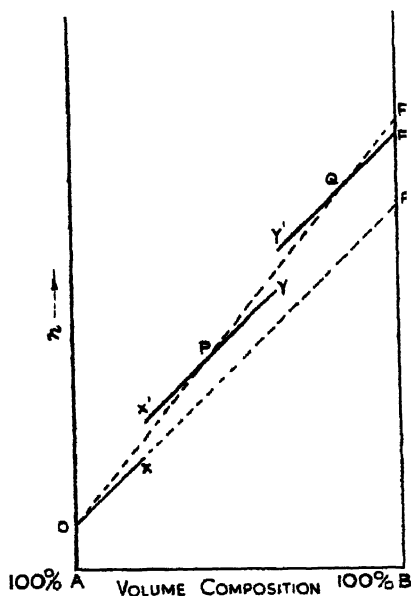
$1\frac{1}{2}$ fringes and the same effect is observed, and so on. The point O corresponds to the refractive index n_A of the gas A and the point F' to n_B .

At points P and Q the dispersion shift is an integral number of wave-bands, and the interference system of the gas tubes is symmetrical and can be brought into exact coincidence with the fiduciary system. The spine O, P, Q....F indicates the balance in non-homogeneous light which would be obtained if the achromatic centres of the fiduciary and gas systems were brought into coincidence, and corresponds to the equation

$$f = (\bar{n}_B - 1)/(\bar{n}_A - 1).$$

The practical balance in which the nearest achromatic band of the gas-tube system is brought into coincidence with the central band of the fiduciary system is given by the family of parallel segments

$$OX, \quad X'Y, \quad \dots\dots Y'F''.$$



A total dispersion shift of 2.4 bands has been shown in the above diagram.

In addition, therefore, to the two values of the density ratio f and f' already considered, there may be added a third value, f'' , which is the one actually observed. In practice it is sometimes difficult to distinguish between f and f' , and in the experiments with liquids already described no attempt was made because of the experimental difficulty.

It may be noted that any observation in white light will agree with monochromatic light if it fall on the first segment OX of the broken line; on other segments there will be a dispersion error of an integral number of wave-bands.

The difference in slope of OF and OF' depends on the difference in dispersivity of the two gases. It happens that the dispersive power of air is low, so that in using it as a comparison fluid, the effect of the dispersion difference is almost always in the sense to make the apparent index of refraction of the second gas or vapour higher than the true value.

Measurements made on mixtures of benzene vapour with air gave a family of curves exactly like those in the figure. It was later found that the same observations had already been made by Kinder (1938), who refers to an important paper by Hansen (1930) in which a formula is given for the location of the *Sprungstellen*.

As already indicated, however, this discontinuity effect is implicit in the earlier treatment by Clack.

§ 7 THE EFFECTIVE WAVE-LENGTH OF THE SOURCE

It is usually assumed that white-light observations correspond to a wave-length of about 0.56μ at the maximum sensitivity of a normal eye. With a tungsten lamp, however, the emission varies rapidly at this wave-length. A rough calculation from specimen curves indicates a visual maximum some 0.02μ higher on this account.

With the writer's right eye (of distinctly different colour sensitivity from the left eye) no difference in band spacing could be seen or measured as between sodium fringes or white-light fringes.

Measurement of the change of air pressure necessary to give a shift of a few fringes showed an effective wave-length of 0.59μ for an ordinary clear half-watt lamp, and 0.585μ for a daylight lamp. It must be admitted, however, that neither of these tests is very satisfactory. So long as a white-light fringe is sufficiently clear to be separated, a sodium fringe will necessarily coincide more or less with the brightest part of it, and in the other case where measurements of fringe shift are attempted, the fringes compared are of reversed colouring. The results in any case refer only to the observer's eye.

With two substances of moderate dispersive power, an error of 0.02μ in the effective wave-length will have little effect on the ratio of the refractivities. It is a point, however, which needs further experiment.

The introduction of strongly coloured filters—tricolour filters for example—has practically no effect on the position of optical balance. The only effect is to trim off the coloured edges from the fringes. It is probable that this would remain true if the spectral width of the source were decreased down to something approaching monochromatic light. On this assumption it is clear that any method based on identifying the central band, even if the final balance were made in monochromatic light, would still yield results governed by group-velocity considerations.

§ 8 THE DISPERSION DETERMINATION

Since the optical balance in white light is a sort of integration result, it would be unwise to expect that refractive indices for values of λ far removed from the effective wave-length of the source could be determined with equal precision. It is possible, however, by careful observation, to determine the extent of the shift of the achromatic band of the gas system and so to find the difference between the dispersion of the gas under examination and that of the reference substance. Since it is the difference which is measured, the total dispersion can be assessed with fair accuracy.

In the observations on gases which follow it has been assumed that the dispersion is sufficiently well expressed by a simple two-term Cauchy expression, $(n-1)=(A+B\lambda^{-2})$, and this is probably true for all colourless gases in the middle of the visual region, indeed the expression fits many liquids and solids surprisingly well.

From the considerations already mentioned, the values of A and B may be determined. It is of interest to note that n_F and n_C may also be obtained with fair accuracy by correcting the value of f' found by the quantity corresponding to the integral dispersion band-shift at the given wave-length.

The extent of the dispersion shift is sometimes very considerable. For example with 100-cm. tubes, in changing from air to ammonia, there is a wandering of the achromatic centre of 19 wave-bands and with air to ethylene about 40 wave-bands. The suggestion appearing in text-books that the central band should be identified in white light and the fractional part determined in monochromatic light is entirely misleading.

§ 9. THE REFRACTION OF AIR

A mean of the determinations by various authors up to about 1930 gave the following values for the refraction of air at N.T.P.:—

λ	0.486 (F)	0.589 (D)	0.656 (C)
$(n-1)10^7$	2947	2926	2916

These values were used in all the present author's work on gases. After this part of the work was finished, it was found that Barrell and Sears, of the N.P.L., had published the results of an elaborate investigation (1939) leading to a value of $(n_D-1)10^7$ at N.T.P. of 2924 and a slightly greater dispersion than is given above. It was not thought necessary to recalculate all the present work on account of these small changes.

§ 10 MEASUREMENTS ON GASES

The interferometer used for the bulk of the work was the laboratory-constructed instrument of the Rayleigh pattern with 50-cm. tubes mentioned above. The light source was an ordinary "coiled-coil" half-watt lamp.

The increase or decrease of pressure in the air tube required to restore optical balance was measured on suitable mercury manometers of 8 mm bore.

The pure gases were made by approved chemical methods and dried, when possible, by a long tube of potassium hydroxide. Measurements were also made on commercial gases from cylinders, and were found to differ only slightly from those on the pure gases.

The dispersion shift in replacing air by the gas under examination was determined by allowing the gas to mix in slowly, and measuring the change in pressure for each half-fringe shift of the achromatic centre, i.e. from a clearly defined single central achromatic band, to the condition where there are two equally coloured bands.

The following table summarizes the mean results obtained for $(n-1)10^7$ at N.T.P.:—

Gas	$\lambda 0.486$	$\lambda 0.589$	$\lambda 0.656$	Dispersion shift
H_2	1408	1394	1388	1.3
	1411	1395	1387	
O_2	2738	2713	2702	2.0
	2743	2714	2701	
N_2	3004	2982	2973	none
	3006	2984	2973	
NH_3	3902	3846	3823	9.5
	3918	3846	3826	
CO_2	4524	4490	4469	1.0
	4528	4492	4473	

The top line of figures gives the observed values for each gas, and the second line shows recent results determined in monochromatic light by other observers.

There appears to be a general tendency for the present method to give a somewhat low result at the blue end of the spectrum. It is likely that this is due to the paucity of blue light in artificial illumination, the refractive index is rising most sharply where visual acuity is rapidly diminishing, and, therefore, the average is somewhat low.

A specimen of the observations and calculations is given for H_2 . In the case of NH_3 , a special correction was necessary from 20° to 0° owing to its abnormal coefficient of expansion.

Specimen observation and calculation for H_2

	cm. Hg	
Barometer	77.09	temp. 20° c.
Air pressure at balance	37.07	
Avogadro correction	0.02	(negligible)
	<u>37.05</u>	

Allowing 1 band dispersion shift = 0.33 cm. pressure,

$$f' = 36.72/77.09 = 0.4763.$$

$$\therefore (n_D - 1)10^7 = 2926 \times 0.4763 = \underline{1394}.$$

For the dispersion :

Total shift of achromatic centre = 13 bands = 0.43 cm. Hg. Hence

$$f = 3715.7709 = 0.4819$$

$$-(\lambda \, dn/d\lambda)_D = 2B\lambda^{-2} = 3017 \times 0.4819 - 1394 = 60.$$

$$\therefore B = 300/2.88 = 10.4.$$

From which

$$n_F = 1394 + 10.4 \times 1.35 = \underline{1408},$$

$$n_C = 1394 - 10.4 \times 0.56 = \underline{1388}.$$

An alternative approximate procedure which entirely ignores the light quality is to use the appropriate pressure corrections for the F and C lines instead of 0.33 cm. as for D above.

This gives

$$n_F = \frac{37.05 - 0.27}{77.09} \times 2947 = 1406,$$

$$n_C = \frac{37.05 - 0.37}{77.09} \times 2916 = 1387.$$

As the pressure measurements are purely relative, it may be noted that there is no need to apply any corrections to the barometer or manometer readings, providing, of course, that these instruments are not far apart.

Using the values of n_D obtained in the present work, and published values for other quantities, the molecular refraction was calculated for the gaseous and liquid state for these gases

For gases, the formula $M_D = 14940(n_D - 1)/(1 + \alpha)$ was used, where $(1 + \alpha) = \text{Avogadro factor or apparent degree of association}$.

Values of M_D

Gas	Gaseous	Liquid
H ₂	2.08	2.09
O ₂	4.05	4.05
N ₂	4.45	4.37
NH ₃	5.65	5.56
CO ₂	6.66	6.57

Considering the difficulties of observation on liquefied gases, the agreement must be regarded as remarkably close.

§ 11 MEASUREMENTS ON AIR-VAPOUR MIXTURES

Mixtures of air with the vapour of organic liquids were made by breaking sealed bulbs of the liquid in a 20-litre bottle which was rotated until the liquid had entirely vaporized. By suitable temperature and pressure measurements, the initial and final volumes were determined, and hence the volume of the vapour found by difference. The mixture was drawn through the tube of the refractometer, which had already been charged with an air-vapour mixture of approximately the same composition in order to minimize sorption effects.

It was found (contrary to usual belief) that the solvent vapours at quite low concentrations (1 to 5 % vol) are appreciably associated, the following values being obtained :—

Methyl alcohol	1.03
Formic acid	1.65
Ethyl alcohol	1.03
Acetic acid	1.84
Acetone	1.01
Methyl ethyl ketone	1.03
Benzene	1.04
Paraldehyde	1.01
<i>n</i> -Heptane	1.04

Vapour refractions are commonly reported at N.T.P. This practice has little to commend it since it involves a most uncertain extrapolation to impossible conditions. It appears more useful to give the specific or molecular refraction and the apparent association factor.

With all the above solvents, the molecular refractions, with the single exception of formic acid, were within 1 % of the liquid values (Lorenz-Lorentz). In the case of formic acid somewhat lower and inconsistent results were obtained, but this appeared to be due to actual chemical attack on the brass refractometer tube.

Attempts were also made to take measurements on mixtures of vapours with hydrogen, but owing to various experimental difficulties this had to be abandoned.

§ 12. TILTING-PLATE COMPENSATORS

The group-velocity or group-index principle may easily be applied to calculate a calibration curve for this type of compensator. Such compensators, however, appear to show quite perceptible changes in calibration in use and are not suitable for work of more than moderate precision.

In some applications, however, they are useful as null indicators, or for measuring small deviations. For example, if it be desired to make a 10 % vol. mixture of a given gas with air, two gas tubes can be used, one 100 cm. long for the mixture and one 10 cm. long for the given gas, and optical balance should result.

This device of a short tube is incidentally, also useful for balancing gases of high refraction by a reasonable pressure of air in the longer comparison tube.

REFERENCES

- ADAMS, L., 1915. *J. Amer. Chem. Soc.* **37**, 1181.
 BARRELL, H. and SEARS, J. E., 1939 *Trans Roy Soc. A*, **238**, 1.
 CLACK, B., 1925 *Proc Phys. Soc.* **37**, 116.
 CRAVEN, E. C., 1939. *Ind Chemist*, **15**, 354.
 GUTTON, C., 1911 *C.R. Acad. Sci., Paris*, **152**, 1089
 HANSEN, G., 1930. *Z InstrumKde*, **50**, 460
 HOUSTOUN, R. A., 1944. *Proc Roy Soc. Edinb A*, **62**, 58.
 KINDER, W., 1938. *Z. Nachr* **1**, 223.
 MANN, C., 1902 *Manual of Advanced Optics*.
 WERNER, G., 1925 *Z. angew. Chem* **38**, 905.
 WILLIAMS, W. E., 1932. *Proc. Phys. Soc.* **44**, 453.

THE INFLUENCE OF ABSORPTION ON THE SHAPES AND POSITIONS OF LINES IN DEBYE-SCHERRER POWDER PHOTOGRAPHS

BY A. TAYLOR AND H. SINCLAIR

Communicated by Sir Lawrence Bragg, F.R.S. ; MS. received 23 October 1944

ABSTRACT Simple graphical methods of determining the basic line-contours for powder diagrams taken in cylindrical Debye-Scherrer cameras are described. From these contours the absorption factor may be obtained for any given Bragg angle and dilution of the specimen. The contours also enable the position of the line peaks to be calculated for different sets of experimental conditions. This, in turn, opens up the possibilities of new types of extrapolation curves for the accurate determination of lattice parameters.

§1 INTRODUCTION

The absorption factor

THE usual experimental arrangement for taking Debye-Scherrer powder photographs is one which employs a cylindrical specimen made of an aggregate of crystal fragments which is placed at the centre of a cylindrical camera. The slit system of the camera is fixed and limits the height of specimen irradiated. In the absence of any absorption of the x-ray beam within the powder aggregate, the intensity of an x-ray reflexion is proportional to the volume of powder irradiated and, therefore, to the area of cross-section of the specimen. In practice there is an appreciable amount of absorption, and the intensity of the reflexion is cut down to a fraction A of the original amount. A is termed the absorption factor. It may be thought of as consisting of two parts, the one being concerned with absorption within the bulk of the powder aggregate (macro-absorption), and the other being a function of absorption and extinction within the reflecting polycrystalline particle (micro-absorption).

We may write the absorption factor in the form

$$A = \tau\alpha, \quad \dots\dots(1)$$

where τ is the micro-absorption factor and α the macro-absorption factor (Schafer, 1933; Brentano, 1935; Brindley and Spiers, 1938; Taylor, 1944). With very small particles ($\sim 10^{-5}$ cm.), τ is nearly 100 %, and the main term in the absorption factor is α . The value of α also has a considerable influence on the distribution of energy in the diffraction line. We shall, accordingly, define the shape of the line as determined solely by α , the basic line-contour, and we shall modify it according to the experimental conditions to obtain the line shapes observed in practice.

The value of the macro-absorption factor α depends on μ , the mean linear absorption coefficient of the powder specimen, and also on r , the radius of the specimen, and on the Bragg angle θ . It is most convenient to compute α in

terms of the product μr for a number of discrete values of θ and to obtain the intermediate values by interpolation. These computations have been carried out analytically by G. Greenwood (1927), A. Rusterholz (1931) for special cases, and graphically by A. Claassen (1930). F. C. Blake (1933) used Claassen's graphical construction with a higher degree of accuracy. The graphical method of Claassen was also given analytical expression by A. J. Bradley (1935), who claims greater accuracy when $\mu r > 2$. L. W. McKeehan (1922) used a very involved analytical method of determining the line shapes for a number of special cases, but did not evaluate the absorption factors from them.

Briefly, and omitting the geometrical constructions involved, the Claassen method is as follows:—In figure 1 the circle represents a cross-section of the cylindrical powder specimen irradiated by a parallel beam of x rays. The total length of path of a ray entering along p and leaving the specimen along q after

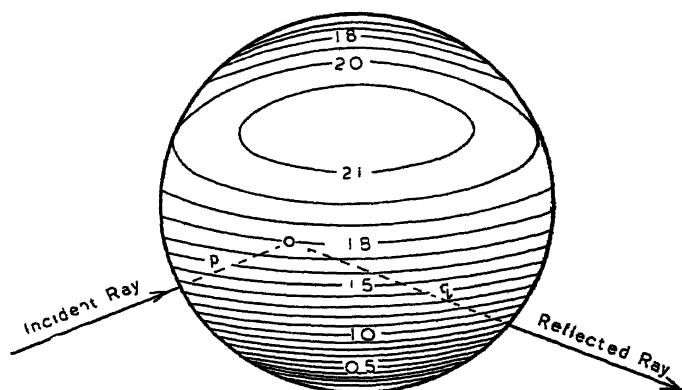


Figure 1. The Claassen loci. Division of specimen cross-section into loci of $z=p+q$ taken at intervals $0.1r$ in z . (After F. C. Blake, *Rev. Mod. Phys.* 5, 169, 1933.)

reflexion from a small element of area $d\sigma$ is $z=p+q$. The macro-absorption factor is given by the expression

$$\alpha = \frac{1}{\pi r^2} \left\{ \int \int e^{-\mu z} d\sigma \right. \quad \left. \dots\dots (2) \right.$$

$$= \frac{1}{\pi} \int \int e^{-\mu r x} ds,$$

where $x=z/r$ and $ds=d\sigma/r^2$. We may treat this final expression as if it referred to a circle of unit radius with an effective linear absorption coefficient μr . We then evaluate the integral for specific values of μr and θ .

In order to evaluate the integral (2), the cross-section of the specimen is divided into loci for which z (or x) is a constant. In figure 1, adapted from Blake's paper, the loci are drawn at intervals of $x=z/r=0.1$. The integration, in the graphical solution, is then carried out by determining the areas between the loci with the aid of a planimeter, multiplying the area of each strip by $e^{-\mu r x}$ and summing the results. The analytical approach achieves the same results by finding the equations to the loci and writing down the values of s and (ds/dx) as a series in x . Bradley gives his results as tables of absolute values of 100α

for different μr and θ values, while Blake plots the ratios of $\alpha(\theta^\circ)/\alpha(90^\circ)$ against θ . This latter form of expression has the disadvantage of yielding only relative values of α .

Now although the analytical and graphical methods described above yield the absorption factor, they do not, in themselves, furnish a simple means of drawing the line shapes. These line shapes are of fundamental importance, for it is essential to know just how far absorption influences the positions of the line peaks in making lattice-parameter determinations, and in the measurement of particle sizes and lattice distortion we must know all the factors which may or may not affect the angular spread of the lines.

We can derive the line shapes directly by means of a very simple strip method, assuming the radiation to be strictly parallel, and from them determine the absorption factors with a high degree of accuracy. In practice, conditions of parallel radiation are not encountered, since the rays diverge from every point in the focal spot. The divergence, as far as the specimen is concerned, is very small, for the angle subtended by the specimen at any point in the focus is only of the order of 40 minutes of arc. The absorption factor is not appreciably influenced by this divergence. However, as we shall see, the line shapes and the peak positions are influenced to a considerable extent by the divergence, for the radius of the camera is comparable with the specimen-focus distance.

To determine the line shapes for divergent radiation, we have to modify the strip method and employ a *line-contour matrix*. This yields the line shape for a point focus, which is then suitably modified to take into account the finite size of the focal spot and the distribution of intensity across it.

§ 2. BASIC LINE-CONTOURS FOR PARALLEL RADIATION

(a) *Line contour and absorption factor by the direct-strip method*

In figure 2, the circle represents a cross-section through the specimen. The parallel beam of radiation travelling in the direction XA completely bathes the specimen, and is diffracted through 2θ in a direction parallel to the central ray AB. The example shown in the figure is drawn for $\theta = 22\frac{1}{2}^\circ$ and $\mu = 2.0$. It is clear that all the energy in the shaded region of the line-contour must have travelled from the shaded strip EF of the specimen. The total energy given

by such a strip must clearly be proportional to $\int e^{-\mu(p+q)} d\sigma$ or to $\int_{\mathbb{R}} e^{-\mu r} ds$ when referred to a unit circle. In practice, it is convenient to make the strips infinitely narrow when the integral taken along a chord such as EF yields the corresponding height of the strip in the line-contour. These integrations are carried out by graphing the function $e^{-\mu r}$ on a large scale and determining the areas by counting squares.

The accuracy with which the integrations yield the heights of the line elements depends, ultimately, on the accuracy with which the lengths p and q can be measured. This accuracy becomes all the more important as the value of μr increases, for then $e^{-\mu r}$ falls rapidly and only the outer skin of the specimen is effective. As long as $\mu r < 5$, a circle 40 cm. in diameter suffices to represent the unit circle on which p and q are measured.

Chords EF were drawn across the unit circle at intervals of 2 cm. In the region of the peak, where additional accuracy was required, the chords were spaced at intervals of $\frac{1}{2}$ cm. Starting at the ends F, where the specimen makes its major contribution to the line intensity and the need for accuracy is greatest, the chords were divided into 1-cm. intervals and the corresponding series of lengths $p+q$ measured off. Smaller subdivisions were taken as required. Tables of $e^{-\mu r}$ were then constructed for each chord EF and for different values of μr and θ , maintaining five-figure accuracy throughout. Thus when the functions were plotted to a large scale for carrying out the integrations, it was possible to obtain an accuracy of at least $\pm \frac{1}{4}\%$ for the ordinates of the line contours.

The area of the basic line-contour is, of course, proportional to the intensity of the diffracted beam, while its outline gives the intensity distribution. When the value of μ is zero, the contour becomes a semi-ellipse of axial ratio 2, with an

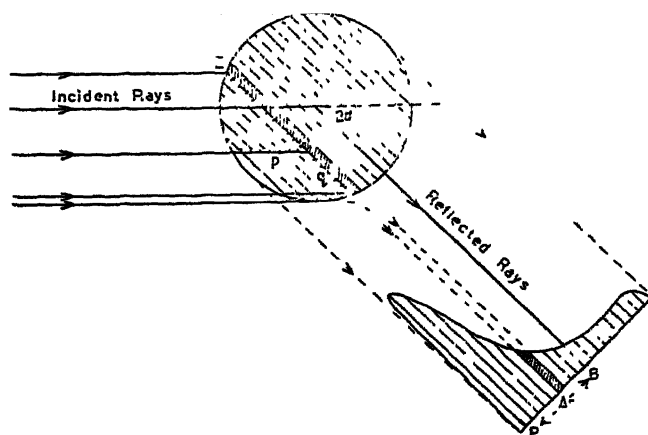


Figure 2. Direct strip method of obtaining line-contour for parallel radiation. Drawn for $\theta = 22\frac{1}{2}^\circ$, $\mu = 2.0$.

area equal to that of the unit circle, π . The absorption factor α for a given value of μr is thus the area under the basic line-contour divided by the factor π .

The method involves a considerable amount of arithmetic and graphical work if high accuracy is aimed at. For the special cases of $\theta = 0^\circ$ and 90° it is possible to cut down the work appreciably without sacrificing any of the accuracy, since the integrations along the chords can be carried out directly.

At 0° , the incident and reflected rays lie in the same straight line, and the total path $p+q$ is equal to the chord length. Thus the total contribution by a given chord of length c of the unit circle is simply equal to $ce^{-\mu r}$. We can thus draw the contour of the diffraction line at 0° with a minimum of error by replacing all measurements of chord length and graphical integrations by direct calculation.

At $\theta = 90^\circ$, the reflected rays travel back along the incident paths, and now $p=q$. The maximum distance traversed by a given ray is now twice the length of the chord whose integrated effect is $\int_0^c e^{-\mu r} \cdot 2x dx$ or $\frac{1}{2\mu r} (1 - e^{-2\mu r c})$. Thus

the ordinates of the line-contour at $\theta = 90^\circ$ may also be computed without the need of intermediate graphical integrations.

A comparison of our final results for the macro-absorption factors at $0, 22\frac{1}{2}, 45, 67\frac{1}{2}$ and 90° with those obtained by Bradley is made in table 1. On the whole the measure of agreement is exceptionally good.

Table 1. Absorption factors 100α when $\mu r < 5$

$\sin^2\theta$ θ°	0.0000 0°		0.1464 22½°		0.5000 45°		0.8536 67½°		1.0000 90°	
μr	Strip method	Bradley	Strip method	Bradley	Strip method	Bradley	Strip method	Bradley	Strip method	Bradley
0.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.25	65.9	65.4	66.2	66.0	65.5	65.5	67.5	67.5	68.0	68.0
0.50	43.0	43.5	44.0	44.2	45.0	44.8	46.8	47.8	48.9	49.0
0.75	28.9	29.0	29.8	29.8	32.8	32.5	35.3	35.4	37.5	37.5
1.00	19.4	19.77	21.0	20.95	23.9	24.2	27.5	27.85	29.5	29.5
2.00	4.71	4.71	6.34	6.35	9.93	10.05	13.8	13.84	15.7	15.67
3.00	1.40	1.436	2.89	2.885	5.75	5.82	8.9	8.89	10.54	10.54
4.00	0.50	0.568	1.71	1.706	3.95	4.02	6.53	6.53	7.94	7.94
5.00	0.26	0.276	1.19	1.189	2.96	3.05	5.14	5.14	6.35	6.35

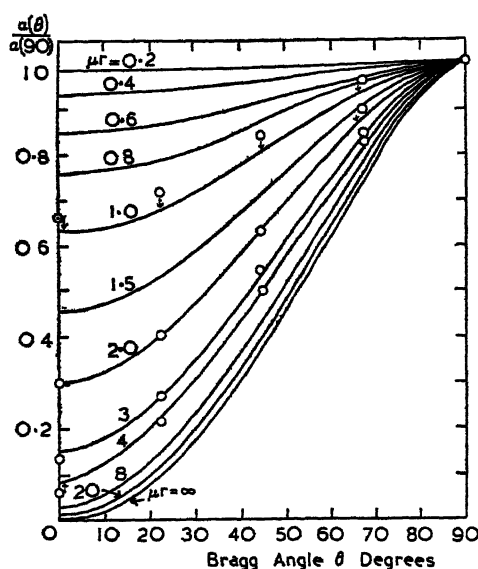


Figure 3. Ratio of absorption factors $\alpha(\theta^\circ)/\alpha(90^\circ)$. (F. C. Blake.)
Circles show present set of results.

We have also expressed our results in the form $\alpha(\theta^\circ)/\alpha(90^\circ)$, and plotted them on Blake's curves, shown in figure 3. The agreement is almost perfect except for $\mu r = 1.0$, when the points come appreciably higher.

The measure of agreement shown between calculations based on the strip method and the values of α given by Blake and Bradley is a verification of the

high accuracy with which the basic line-contours have been drawn. A set of these contours is illustrated in figure 4. For convenience of representation they have all been drawn the same height. The basal widths are on the same scale, being equal to $2r$, the diameter of the specimen. When they appear to be much narrower, as, for example, when $\theta = 45^\circ$ and $\mu r = 4.0$, it is because the intensity falls steeply and the vertical scale is too small to reveal the weak tails on the lines. The numbers uppermost on the curves are the corresponding values of 100α . These are proportional to the areas beneath the contours and give some idea of the magnification employed in plotting the contours of lines with high values of μr .

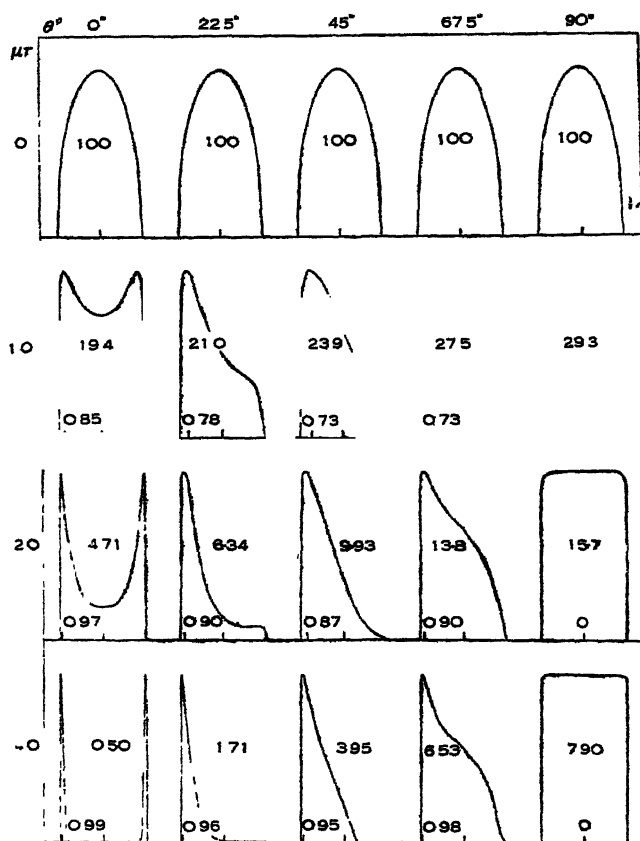


Figure 4. Basic line shapes of Debye-Scherrer lines with parallel radiation obtained with the direct-strip method.

(b) Position of line peaks with parallel radiation

The peaks of the lines are displaced from the central ray in the direction of greater θ by a distance Δr along the film. For parallel radiation and a given value of θ , Δr will be proportional to the radius of the specimen, provided μr is maintained constant. Thus it is convenient to express the peak displacement as a fraction of the radius $\Delta r/r$, and, since parallel radiation is employed, this fraction must be the same for all values of the camera radius R . Values of $\Delta r/r$ are given at the base of the lines in figure 4.

For a line having a flat top, we have adopted the convention of referring $\Delta r/r$ to the centre of the flat portion, as, for example, at $\theta=90^\circ$. When the line displays a definite peak, the measurement is taken at the point of maximum intensity. $\theta=0^\circ$ represents a special case. Here the peak is at the centre for

Table 2. Values of $\Delta r/r$ for parallel radiation

μr	$\theta=0^\circ$	$22\frac{1}{2}^\circ$	45°	$67\frac{1}{2}^\circ$	85°	90°
0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.25	0.00	0.28	0.30	0.24	0.07 _s	0.00
0.50	0.00	0.58	0.47	0.48	0.22	0.00
0.75	0.72 _s	0.72 _s	0.64	0.64	0.50	0.00
1.00	0.85	0.78	0.72 _s	0.73	0.67 _s	0.00
2.00	0.97	0.90	0.87	0.90	0.90	0.00
3.00	0.98	0.94	0.92	0.96	0.97	0.00
4.00	0.98 _s	0.96	0.95	0.97 _s	1.00	0.00
5.00	0.99 _s	0.97 _s	0.97 _s	0.98 _s	1.00	0.00
∞	1.00	1.00	1.00	1.00	1.00	0.00

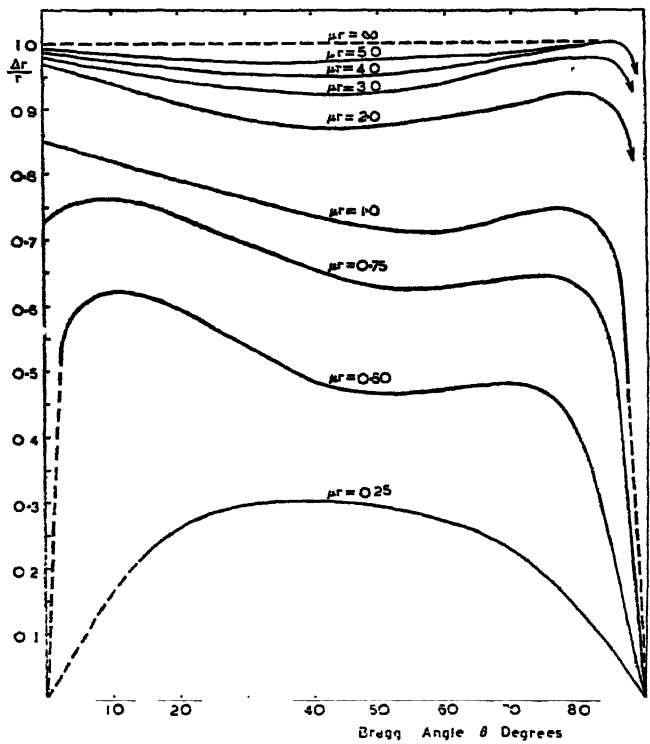


Figure 5. Values of $\Delta r/r$ as a function of μr and Bragg angle θ Calculated for parallel radiation.

$\mu r=0$. As μr increases, the contour flattens out and then becomes slightly depressed at $\mu r \simeq 0.5$. The line suddenly switches over from having one peak to having two, which lie close to the outer edges of the line. Near the value of

$\mu r = 0.5$, when the change-over is on the point of taking place, the position of the peak is indeterminate.

Values of $\Delta r/r$ are given in table 2. Between $67\frac{1}{2}^\circ$ and 90° the values fall rapidly to zero. This angular region is of particular importance in lattice-parameter determinations, for the resolving power of the powder diagram increases as $\tan \theta$. The maximum value of θ recorded in cylindrical cameras rarely exceeds 86° . The line-contours for $\theta = 85^\circ$ were therefore calculated, and values of $\Delta r/r$ obtained from them to bridge the gap between $67\frac{1}{2}^\circ$ and 90° . These values are also recorded in the table.

Figure 5 illustrates the remarkable variations of $\Delta r/r$ much more clearly than the table. At angles below $22\frac{1}{2}^\circ$ the positions of the peaks for $\mu r < 0.75$ are difficult to determine even with very accurately drawn contours. Between 80° and 90° the curves suddenly fall to zero and make interpolation in this region very difficult.

§3 BASIC LINE-CONTOURS FOR RADIATION DIVERGING FROM A POINT SOURCE

When we consider the line-contours for divergent radiation, we find that, except for the special cases of $\theta = 0^\circ$ and 90° , they cannot be derived from the contours for parallel radiation by any simple geometrical construction. The contours of the lines now become dependent on the camera radius R , and the specimen-focus distance AX in addition to θ and μr . We shall find that as long as the focus may be regarded as a point source, the dimensions of the line-contour will change in proportion to r provided μr is maintained at the same value and the other variables are unchanged. When the focus becomes finite, the proportions of the line-contour then change according to the ratio : size of focal spot/radius of specimen.

In our analysis we have considered two sizes of camera, namely, cylindrical cameras of radius $R = 950$ mm. and 4500 mm. For these cameras, specimen-focus distances of 150 mm. and 100 mm. respectively have been taken as values most nearly realized in practice.

In figure 6, X represents the point source, radiation from which completely bathes the specimen. We may divide the specimen into strips EF , which subtend equal angles at the point X . Since, for a specimen of 1 mm. radius at a distance $AX = 150$ mm., the total angle subtended by the specimen is only $2/150$ radian, or 0.76° approximately, these narrow strips may be considered as rectangles tilted slightly with respect to each other.

The rays diffracted from the strip EF through an angle 2θ will strike the film over the range GH with an intensity distribution shown schematically by the shaded area. This distribution is easily obtained from the tabulated values of $e^{-\mu(\varphi + \vartheta)}$ obtained for parallel radiation. We then have to determine the relative positions on the film of the contributions from strips such as EF . This is done by calculating the exact location of points G with respect to B , the point where the central ray, unaffected by divergence, strikes the film. The overlapping contributions from EF and from the central strip are shown (not to scale) in the figure.

To compute the ordinates which yield the line-contour, we proceed as follows. We divide the basic circle into twenty parallel strips and plot graphs of the functions $e^{-u(p-q)}$ for each of the strips. (These were already available from the previous work for parallel radiation.) Starting with the graph for the outermost strip near the peak, we read off values of $e^{-u(p+q)}$ at set intervals and write these down in a row. We then calculate the position of G, the origin of the next strip, and commence reading values from the appropriate curve of $e^{-u(p+q)}$ at positions which coincide with the intervals in the preceding strip. These numbers form the next row of figures. Proceeding in this manner for all the chords EF, we build up the *line-contour matrix*. Such a matrix for $R = 45.0$ mm, $AX = 100.0$ mm. and $\mu r = 2.0$ for $\theta = 45^\circ$ is shown in figure 7.

The figures along each horizontal row of the matrix correspond to the energy contributions given by each strip EF, and the envelope of the matrix corresponds to the relative positions of the terminal points GH of the strip contributions. The ordinates of the basic line-contour are then obtained by graphing the numbers

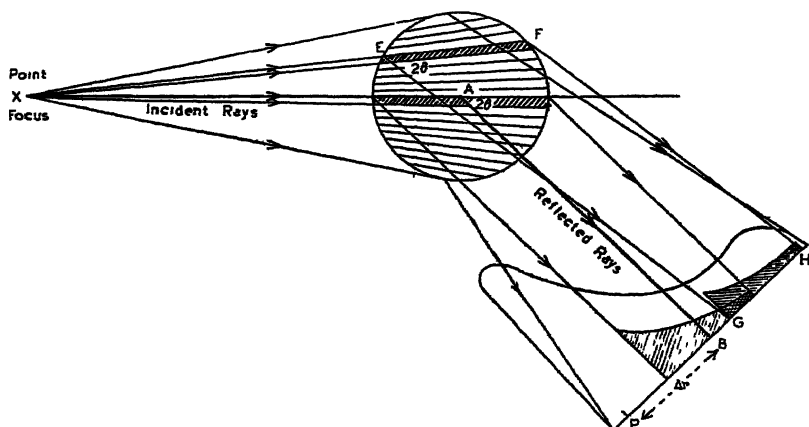


Figure 6 Construction of basic line-contour for divergent radiation

in the vertical columns and finding the areas beneath the curves. This yields a much greater degree of accuracy than is to be obtained by simply adding the numbers together. In the region of the peak, the vertical columns are taken at $\frac{1}{2}$ and $\frac{1}{4}$ intervals in order to determine the peak positions with greater accuracy. For the sake of clarity, these extra columns have been omitted from the diagram.

In figure 8 we give a comparison of line-contours for parallel radiation and for a point focus when $R = 95.0$ mm., $AX = 150$ mm., and $\mu r = 2.0$. For convenience of reproduction they are all drawn to the same maximum height. At angles below $\theta = 45^\circ$ the lines from the point focus are broadened, while above 45° there is a marked focusing effect and the lines are sharpened. At 0° the horizontal dimensions are increased by a factor $\left(1 + \frac{R}{AX}\right)$ while at 90° they are reduced in the ratio $\left(1 - \frac{R}{AX}\right)$. Despite these dimensional changes, the final areas under the contours, and, therefore, the absorption factors, must remain unchanged,

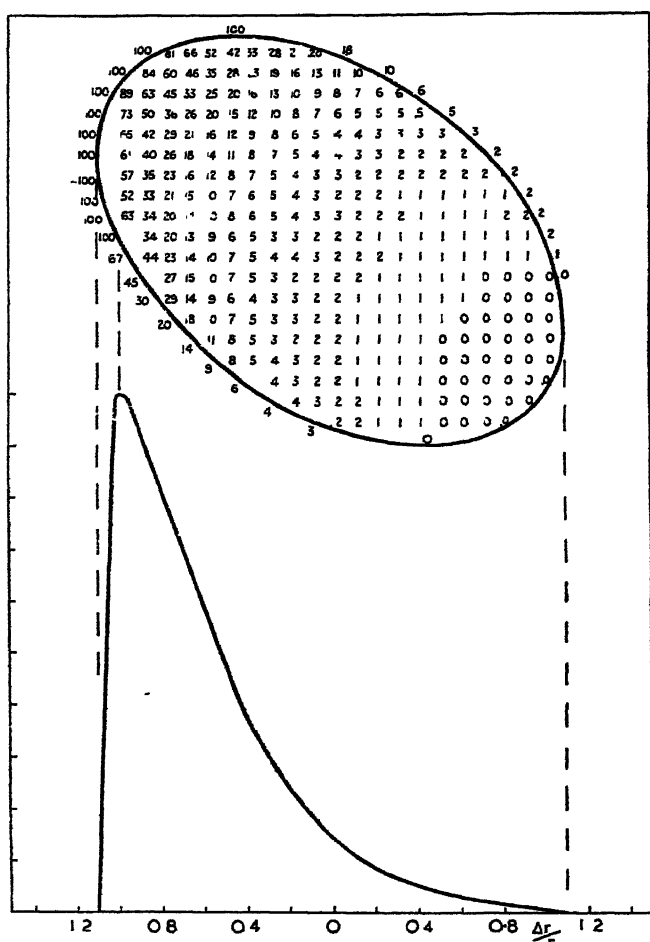


Figure 7. Generation of line-contour by means of a contour-matrix.
 Drawn for $\theta=45^\circ$, $\mu r=2.0$, $R=45.0$ mm., $AX=100.0$ mm. and for point focus.

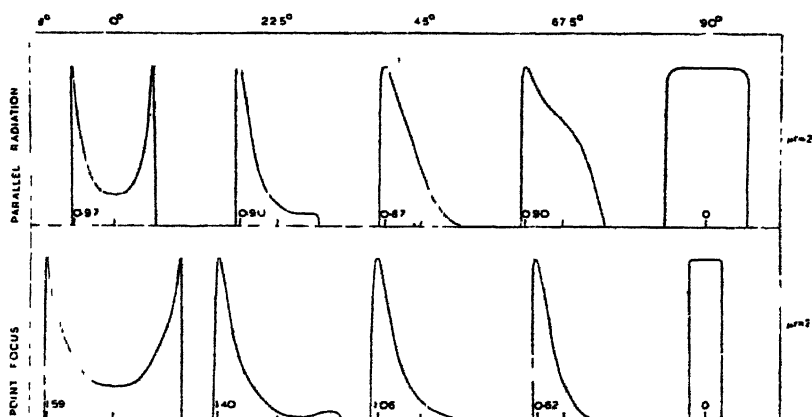


Figure 8. Comparison of basic line-contours for parallel radiation and for point focus.
 Drawn for $AX=150.0$ mm., $R=95.0$ mm., $\mu r=2.0$.

for the summations for divergent radiation must be the same as for parallel radiation. All we have done is to change the order in which the summations are carried out. This result agrees with the analytical expression of Rusterholz (1931) when the divergence is small.

If we increase the radius r of the specimen, we increase the lengths of the rows in the matrix in proportion. We also increase the angle subtended by the

Table 3. Values of $\Delta r/r$ for divergent radiation

Small camera of radius $R=45.0$ mm.

Focus-specimen distance $AX=100.0$ mm.

$\mu r \backslash \begin{matrix} \sin^2 \theta \\ \theta \end{matrix}$	0 0°	0.1464 22½°	0.5000 45°	0.8536 67½°	1.0000 90°
0.00	0.00	0.00	0.00	0.00	0.00
0.25	0.00	0.36	0.30	0.15 ₅	0.00
0.50	0.00	0.69 ₅	0.55	0.28	0.00
0.75	1.05 ₀	0.95 ₅	0.72	0.38	0.00
1.00	1.23 ₂	1.09	0.83 ₅	0.46	0.00
2.00	1.40 ₉	1.25	1.00	0.63	0.00
3.00	1.42	1.30	1.02 ₇	0.71	0.00
4.00	1.43 ₅	1.33	1.05	0.73	0.00
5.00	1.44	1.33 ₆	1.06	0.73 ₅	0.00
∞	1.45 ₀	1.34 ₅	1.09 ₈	0.75 ₀	0.00

Table 4. Values of $\Delta r/r$ for divergent radiation

Large camera of radius $R=95.0$ mm

Focus-specimen distance $AX=150.0$ mm.

$\mu r \backslash \begin{matrix} \sin^2 \theta \\ \theta \end{matrix}$	0 0	0.1464 22½°	0.5000 45°	0.8536 67½°	1.0000 90°
0.00	0.00	0.00	0.00	0.00	0.00
0.25	0.00	0.54	0.44	0.23	0.00
0.50	0.00	0.96	0.76	0.41	0.00
0.75	1.18 ₅	1.81 ₁	0.90	0.49 ₅	0.00
1.00	1.39	1.28	0.97	0.55	0.00
2.00	1.58 ₅	1.40	1.06	0.62	0.00
3.00	1.60	1.45	1.07 ₅	0.65 ₅	0.00
4.00	1.61 ₅	1.46	1.12	0.66 ₅	0.00
5.00	1.62 ₂	1.47	1.13	0.68	0.00
∞	1.63 ₅	1.51	1.18 ₅	0.71 ₁	0.00

specimen in the same ratio, and so the matrix rows move a proportional amount relative to each other. In other words, it is as if the matrix were drawn out on a sheet of elastic which can be stretched horizontally in proportion to the radius of the specimen. If, therefore, we keep μr constant while at the same time increasing r , by diluting the specimen, the numbers in the matrix remain exactly the same, and the line-contour retains the same relative proportions. Thus

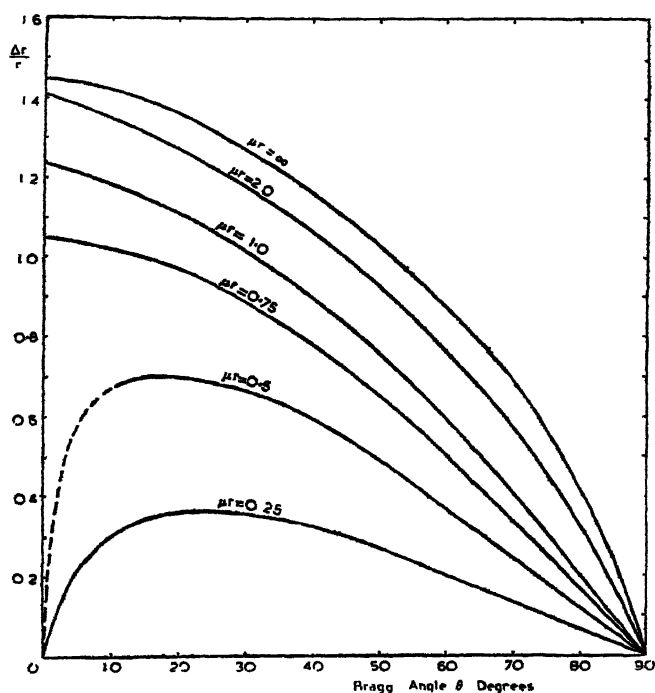


Figure 9 Peak displacement of curves $\Delta r/r$ for point focus.
For $R=45.0$ mm, $AX=100.0$ mm.

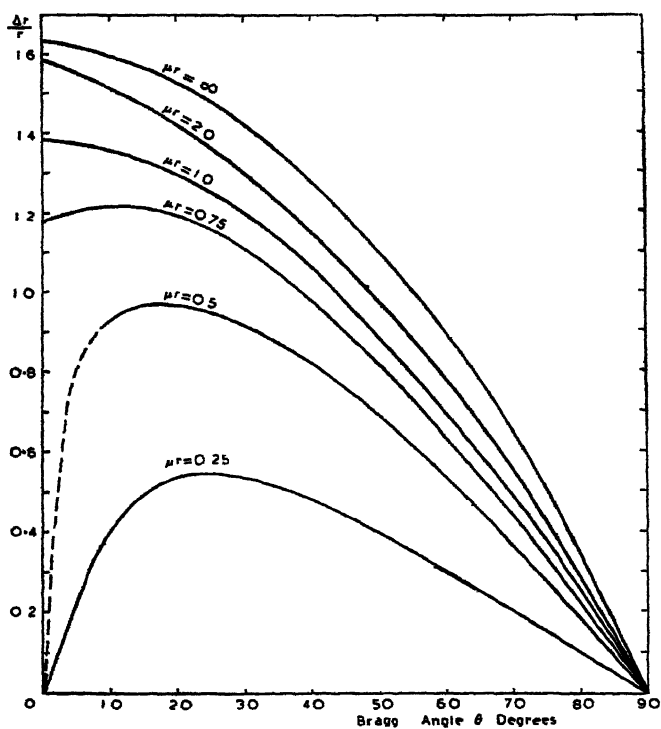


Figure 10. Peak displacement curves $\Delta r/r$ for point focus.
For $R=95.0$ mm., $AX=150.0$ mm.

the displacement of the peak Δr from the central ray PB in figure 6 is directly proportional to the radius of the specimen. Thus the ratio $\Delta r/r$ is constant for a given set of conditions for all values of the radius r , provided μr is kept the same. This breaks down in the case of a finite focus.

Values of $\Delta r/r$ for divergent radiation for two standard camera sizes are given in tables 3 and 4.

The variations in $\Delta r/r$ with μ for different values of θ are best followed from figures 9 and 10. The run of these curves is totally different from those shown in figure 5 for parallel radiation, being almost linear in the range from $\theta = 45^\circ$ to $\theta = 90^\circ$. For $\mu r < 0.75$ and $\theta < 22\frac{1}{2}^\circ$ the curves drop suddenly to the origin. The precise positions of the peaks in this region are difficult to determine with accuracy, and the $\Delta r/r$ curves are drawn dotted to suggest their probable courses. When μr becomes greater than 1.0, the curves of $\Delta r/r$ crowd together very rapidly on the limiting curve for $\mu r = \infty$. Values of $\Delta r/r$ for $\mu r = \infty$ are easily determined, for they coincide with the high-angle outer edge of the line-contour matrices.

§4. MODIFIED LINE-CONTOURS

In the above derivations of the basic line-contours we have chosen a set of highly idealized conditions. In practice, neither a truly point source nor an exactly parallel beam of x rays is encountered. Moreover, the monochromatic beam we tacitly assumed to exist occupies in reality a definite, though very narrow, wave-band. We have also assumed the reflecting crystals to be perfect and to reflect the incident radiation over a negligibly small range of θ . When the crystals are very small ($\sim 10^{-5}$ cm.) or distorted, there is an additional angular spread which may be many degrees, corresponding to several millimetres spread on the powder diagram. Finally, when the line is microphotometered, we have to consider the effect of the finite slit in the instrument and the characteristic curve of the film emulsion. All we can do here is to indicate which factors are operative, and we shall limit ourselves to the effects produced by the focal spot and the photometer slit upon the basic line shapes.

(a) *Effect of focal spot*

In the majority of crystallographic tubes, the focal spot is in the form of a horizontal bar with approximate dimensions 8.0×1.0 mm. The beam is taken off at an angle of 5° to 10° , thus foreshortening the focus to an effective area $\simeq 1$ mm. square. Treating each point in the focus as a source of divergent rays, each point will give rise to a basic line-contour slightly displaced from its neighbour on the x-ray film. The final line shape is the integrated effect of all these contours.

We shall consider the case of $R = 95.0$ mm. and $AX = 150$ mm. A focus with projected area 1 mm. square corresponds to a movement of a basic line along the film of $95 \times 1/150 = 0.633$ mm. For simplicity of calculation we shall assume the focus to be slightly smaller than this 1 mm. and make the movement across the film 0.60 mm.

We shall first take the case of a focus with a uniform distribution of energy. The basic line will move through a distance of ± 0.3 mm. on each side of its mean position as shown in figure 11 (*d*), drawn for $\mu r = 2.0$, $r = 0.25$ mm., $\theta = 22\frac{1}{2}^\circ$

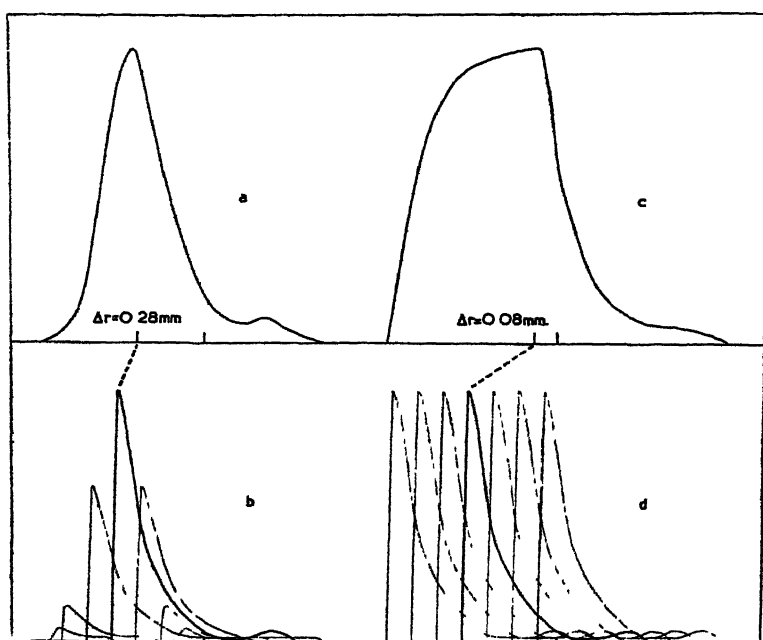


Figure 11 Generation of line-contours for uniform and exponential foci.
For $R=95.0$ mm., $AX=150$ mm., $r=0.25$ mm., $\mu r=2.0$, $\theta=22\frac{1}{2}^\circ$.
(a) Line from exponential focus to form e^{-l^2/x^2}
(c) Line from uniform focus 1 mm. across

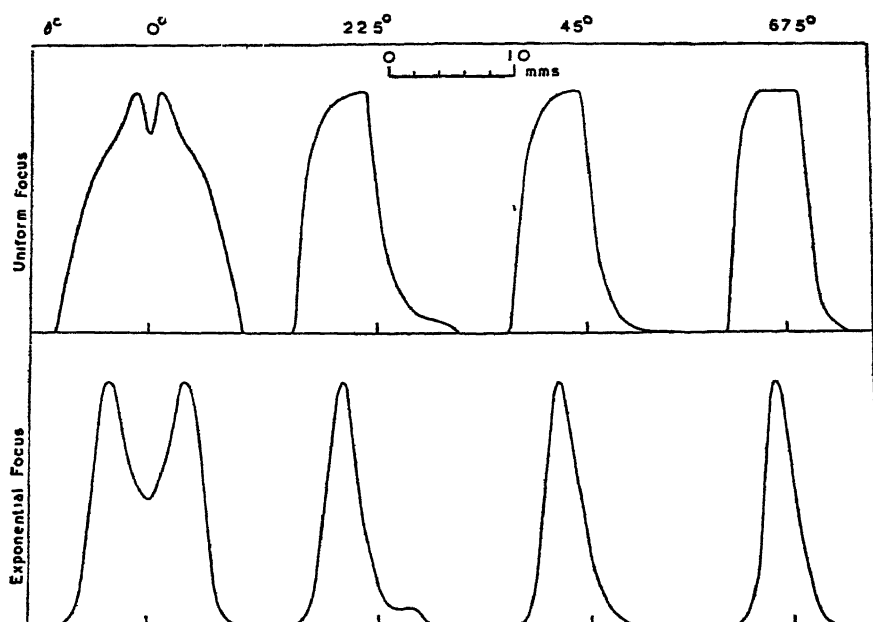


Figure 12. Effect of finite focus on line-contours.
For $R=95.0$ mm., $AX=150$ mm., $r=0.25$ mm., $\mu r=2.0$.

The net effect of the movement is the line-contour shown above it in figure 11 (c). The ordinates of the final contour were obtained by forming a matrix having 25 horizontal rows, each row being displaced through an amount equal to 0.60/24 mm. with respect to its neighbour. The final line shape is now quite different in appearance from the basic line-contour, and the displacement of the peak from the central ray is also very much smaller. Had the focus been broader still, the top of the line would have developed a flat central portion.

In figure 11 (a) we illustrate the effect of an energy distribution of the form $e^{-k^2 r^2}$ across the focal spot. Some of the components of the line at $\theta = 22\frac{1}{2}^\circ$ for $\mu r = 2.0$ and $r = 0.25$ mm. are shown in figure 11 (b) to illustrate the build-up,

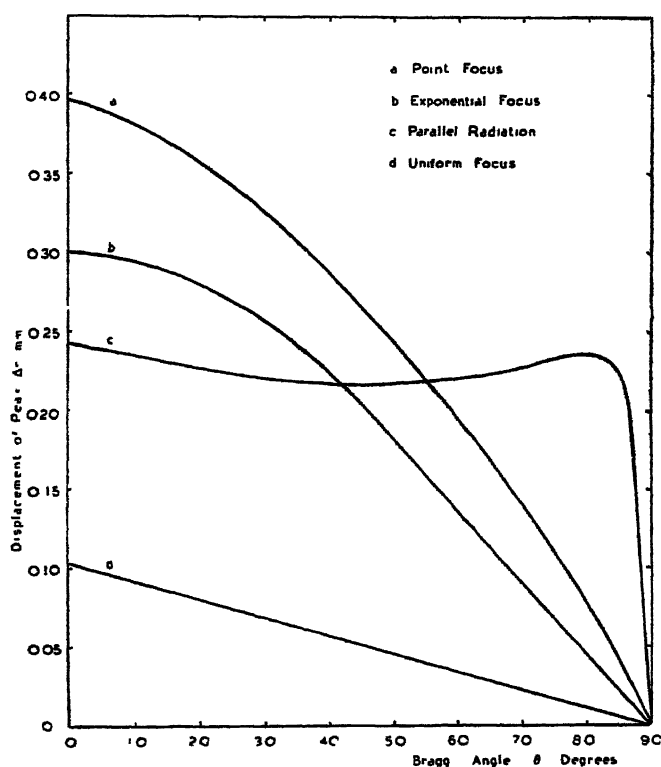


Figure 13 Values of peak displacement Δr mm for different types of focus. For $R=95.0$ mm., $AX=150.0$ mm., $r=0.25$ mm. and $\mu r=2.0$.

which was carried out, as in the previous example, by forming a 25-row line-contour matrix. This line shape corresponds quite closely to the shape found in practice.

A comparison of the line shapes when $AX=150$ mm., $R=95$ mm., $\mu r=2.0$ and $r=0.25$ obtained for uniform and exponential foci is made in figure 12. The effect of crystal imperfection and the finite width of the wave-band employed will be to broaden out the lines still further. These corrections have not been applied as they vary from case to case.

In table 5 is shown how the experimental conditions influence the peak positions. We see that a uniform focus produces least movement of the peak from

the central ray, but the line shape is bad for satisfactory resolution of the spectra. Parallel radiation gives the sharpest peaks, but the lines are very asymmetrical. The line shapes for the exponential focus are by far the most symmetrical, and their well-defined peaks suffer rather less displacement than those for the point focus. The line shapes also correspond most closely to those found in practice.

Table 5. Values of Δr (mm.) for different types of focus

For $R=95.0$ mm., $AX=150$ mm., $r=0.25$ mm., $\mu r=2.0$.

Radiation	$\theta=0^\circ$	$22\frac{1}{2}^\circ$	45°	$67\frac{1}{2}^\circ$	90°
Parallel	0.243	0.225	0.218	0.225	0.000
Point source	0.396	0.350	0.265	0.155	0.000
Uniform source	0.105	0.080	0.050	0.030	0.000
Exponential source	0.300	0.275	0.205	0.100	0.000

Plots of Δr against Bragg angle θ are shown in figure 13. Curve (d) for the uniform focus is interesting, for it reveals an almost perfectly linear relation between Δr and θ for the conditions set forth in the calculation. Curves (a) and (b) for point and exponential foci are almost linear from 45° to 90° .

(b) Effect of finite microphotometer slit

When measuring the intensities of the spectrum lines, or their half-peak widths, a microphotometer is generally employed. The density of photographic blackening B is a measure of the x-ray intensity incident on that portion of the film where the measurement is made. For the relatively high densities obtained in x-ray diffraction work an absorption microphotometer is used. In such an instrument, the image of an incandescent source is focused on the film and the transmitted rays of light are then focused on to a narrow slit stationed in front of a photo-electric cell which registers the amount of transmitted light. The effective slit-width usually corresponds to 0.10 mm. of film, and the film is moved in steps of 0.10 mm., taking readings in each position. These readings, translated into units of photographic density B , give a curve representative of the energy distribution in the spectrum lines. For Debye-Scherrer work the Dobson type of microphotometer is perhaps the most satisfactory, as it yields values of B directly (Taylor, 1945).

If the slit is made too wide, resolving power is lost, whereas if it is made too narrow, the grain in the emulsion tends to be exaggerated. These practical considerations restrict the dimensions of the photometer slit within well-defined limits, and we must see what effect it has on the line shapes.

If the intensity of the light beam focused on the point x of density B_x is i_0 , the amount i transmitted through the film is given by the expression $i=i_0e^{-B_x}$. If the finite slit lies with its centre at x , the instrument reading will only be an average value \bar{B} if there is a density gradient across it. It can easily be shown that $\bar{B}=\log_e \left(1/\int e^{-B_x} dx \right)$, where the integration (performed graphically) is taken over the slit width.

In figure 14 we have chosen the line-contours for $\theta = 45^\circ$ and $\theta = 22\frac{1}{2}^\circ$ to illustrate the effect of a slit 0.10 mm. wide. We find that the lines are slightly broadened at the base while the steeply rising portions are appreciably reduced in intensity. This has the effect of reducing the peak height and reducing the line width without affecting the peak position in curve A for $22\frac{1}{2}^\circ$, while in B, where the density falls sharply on one side of the maximum, there is a marked rounding-off of the peak, and the peak position is displaced away from the steep portion of the curve.

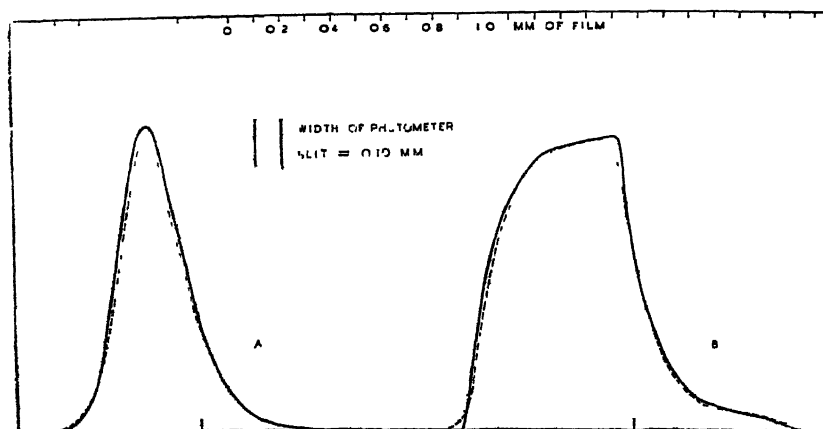


Figure 14. Effect of photometer slit on line shape.

A. $\theta = 45^\circ$, $\mu r = 2.0$, exponential focus

B. $\theta = 22\frac{1}{2}^\circ$, $\mu r = 2.0$, uniform focus.

$R = 95.0$ mm., $AX = 150.0$ mm., $r = 0.25$ mm

Provided that the photometer slit is not too wide in relation to the width of the line, its effect on line shape will be relatively small. Assuming the crystals themselves do not produce an excessive amount of broadening, due to lattice distortion or small grain size, the distribution of the energy in the focal spot will produce the most important modifications in the basic line-contours. This energy distribution is most easily obtained by taking a "pin-hole" photograph of the focal spot.

§ 5. THE INFLUENCE OF LINE SHAPE UPON THE DETERMINATION OF LATTICE PARAMETERS

We have seen that absorption tends to move the peaks in a direction of increasing θ by amounts depending on the experimental conditions. This, in turn, considerably influences the value of the lattice parameters computed from the corresponding θ values of the peaks. To eliminate the errors due to absorption and the displacement of the specimen from the geometrical centre of the camera, various extrapolation methods have been devised which make full use of the high-angle reflexions for which the resolving power of the powder photograph is

Briefly, values of the lattice parameter a obtained from the observed value of θ are plotted against corresponding values of θ , $\cos^2 \theta$, $\cot \theta$ or $(\pi/2 - \theta) \cot \theta$.

The value obtained by extrapolation against the selected function of $\theta=90^\circ$ is substantially free from error. A small correction, approximately one part in 50,000, is added to the parameter to correct for refractivity.

Most generally used is the plot of parameter against $\cos^2 \theta$ with extrapolation to $\cos^2 \theta=0$. Considerations of the curves for $\Delta r/r$ show that plot to be eminently suitable for lattice-parameter determinations when there are sufficient reflexions above 45° . Further consideration shows that if the parameters are plotted against $\frac{1}{2} \left(\frac{\cos^2 \theta}{\theta} + \frac{\cos^2 \theta}{\sin \theta} \right)$, a perfectly straight line plot is obtained from $\theta=0^\circ$ to $\theta=90^\circ$, provided the focus takes the form $e^{-k^2 x^2}$ and the specimen is accurately centred in the camera.* In this case one high-order reflexion in the region of $\theta=70$ to 80° and a very low-order reflexion, $\theta \simeq 10^\circ$, are sufficient to obtain the very highest accuracy in parameter. This important result, and other extrapolation possibilities, are discussed in a later communication.

§ 6. CONCLUSIONS

It has been shown possible to draw the contours of Debye-Scherrer lines for conditions of parallel and divergent radiation, and to obtain from them the macro-absorption factors α with a high degree of accuracy. The distribution of energy in the focal spot was shown to exert a considerable influence on the ultimate line shape, while the effect of the photometer slit was relatively small, though by no means negligible. The accurate determination of peak displacements leads to new forms of lattice-parameter extrapolation curves.

ACKNOWLEDGMENT

The authors wish to thank the English Electric Co., Ltd., for permission to publish this paper.

* The authors are much indebted to Dr. D. P. Raley, of the Cavendish Laboratory, Cambridge, who first used this expression, and for which a theoretical basis can now be given

REFERENCES

- BLAKE, F. C., 1933 *Rev. Mod. Phys.* **5**, 169.
 BRADLEY, A. J., 1935. *Proc. Phys. Soc.* **47**, 879.
 BRADLEY, A. J. and JAY, A. H., 1932. *Proc. Phys. Soc.* **44**, 563.
 BRENTANO, J., 1935. *Proc. Phys. Soc.* **47**, 932.
 BRINDLEY, G. W., *Phil. Mag.* (In the press.)
 BRINDLEY, G. W. and SPIERS, F. W., 1938. *Proc. Phys. Soc.* **50**, 17.
 CLAASSEN, A., 1930. *Phil. Mag.* **9**, 57.
 GREENWOOD, G., 1927. *Phil. Mag.* **3**, 963.
 MCKEEHAN, L. W., 1922. *J. Franklin Inst.* **193**, 231.
 RUSTERHOLZ, A., 1931. *Helv. phys. Acta*, **4**, 68.
 SCHAFER, K., 1933. *Z. Phys.* **86**, 738.
 TAYLOR, A., 1944. *Phil. Mag.* **35**, 215.
 TAYLOR, A., 1945. *Introduction to X-ray Metallography*, p. 92. (Chapman and Hall.)

ON THE DETERMINATION OF LATTICE PARAMETERS BY THE DEBYE-SCHERRER METHOD

BY A. TAYLOR AND H. SINCLAIR

Communicated by Sir Laurence Bragg, F R S ; MS. received 23 October 1944

ABSTRACT The types of systematic error arising in the determination of lattice parameters by the use of Debye-Scherrer powder diagrams are discussed. The various extrapolation methods are reviewed, and it is shown how a consideration of the absorption factor and geometry of the focal spot lead to the most satisfactory forms of extrapolation curve. It is shown how the absence of specimen eccentricity enables perfectly linear extrapolation curves to be drawn, thus allowing fullest use to be made of low-angle reflexions whereby the very highest accuracy in parameter determination may be achieved.

§1. THE DETERMINATION OF LATTICE PARAMETERS

CIRCULAR Debye-Scherrer powder cameras are most conveniently used in the identification of compounds and alloy phases. When the structures have relatively high symmetry, powder photographs may also be employed for the elucidation of their atomic arrangements. At some stage of the work, particularly in the case of alloys when the course of a phase boundary is being followed, it becomes necessary to make precision measurements on the dimensions of the unit cell. If the Debye-Scherrer camera has been satisfactorily designed, it is possible to record diffraction spectra at Bragg angles in the region of 85 to 86° . It is then possible to achieve an accuracy in lattice-parameter measurement comparable with that yielded by flat-film back-reflexion cameras recording only a limited portion of the diffraction pattern.

All parameter determinations by x-ray methods involve the derivation of the Bragg angle θ for a given set of reflecting planes from measurements on the peak positions of the diffraction spectra. These peak positions are very sensitive to the experimental conditions which introduce various systematic errors into the parameter determinations. In our discussion, we shall consider only cubic crystals, and thus confine ourselves to one lattice parameter, although the methods under consideration can, in many cases, be applied to more complicated unit cells.

For a cubic crystal, the lattice parameter a is related to the wave-length within the crystal of the radiation λ , the Bragg angle θ , and the indices of reflexion hkl , by the equation

$$a = \frac{\lambda}{2} \frac{\sqrt{h^2 + k^2 + l^2}}{\sin \theta}. \quad \dots\dots(1)$$

The fractional error in a , caused by errors in measuring θ , is obtained by differentiating equation (1), and is

$$\frac{da}{a} = -\cot \theta d\theta. \quad \dots\dots(2)$$

Thus in the region where θ approaches 90° , $\cot \theta$ tends to zero, and any error in measuring θ produces only a small variation in the lattice parameter. In aiming at the highest accuracy in parameter determination, it is thus essential to design a camera which will record as high a Bragg angle as possible, and to choose a radiation which will give spectra at a high Bragg angle.

If R is the camera radius and S the distance between corresponding reflexions on each side of the incident ray,

$$S = 4R\theta, \quad \text{so that} \quad dS = 4Rd\theta.$$

Hence

$$\frac{da}{a} = -\frac{\cot \theta}{4R} dS. \quad \dots\dots(3)$$

The errors dS which arise from the experimental conditions may be classified under the following headings :—

- (a) Finite length of specimen irradiated by the beam
- (b) Film shrinkage.
- (c) Refractive index of the crystal for x rays
- (d) Eccentricity of the specimen.
- (e) Absorption of the beam within the specimen.

A. J. Bradley and A. H. Jay (1932) have shown that the error in lattice parameter caused by (a) is negligibly small. They eliminate the effects of (uniform) film shrinkage by fitting the camera with knife-edges which subtend a standard angle θ_k in the region of $\theta = 90^\circ$. θ_k is obtained by measuring the camera directly or by taking a photograph of a standard material such as rock-salt or quartz for which the angles of reflexion are known to a very high accuracy (Bradley and Jay, 1933; Wilson and Lipson, 1941). The angle θ of any pair of lines on the film is related to S , the distance between the lines, by the relation $\theta/\theta_k = S/S_k$, where S_k is the distance apart of the fiducial marks on the film produced by the knife-edge shadows.

The correction for refractive index is very small, being of the order of one part in 50,000. This is added to the final value of the lattice parameter (Weigle, 1934; Jette and Foote, 1935).

The errors in lattice parameter introduced by the eccentricity of the specimen and by absorption are most easily eliminated by extrapolation methods making use of high-angle reflexions where the resolving power of the powder photograph is greatest and errors in θ least. The simplest of these methods was described by G. Kettman (1929), who plotted values of lattice parameter a calculated for each line on the film against corresponding values of Bragg angle θ . A smooth curve was drawn through the points, and by extrapolating to $\theta = 90^\circ$, a value of a was obtained largely freed from experimental error. A similar process is described by Bradley and Jay, who plot values of a against the corresponding values of $\cos^2 \theta$ and extrapolate to $\cos^2 \theta = 0$. Extrapolations may also be carried out against $\cot \theta$ or $(\pi/2 - \theta) \cot \theta$, as pointed out by Buerger (1942).

All these methods of graphical extrapolation yield slightly different values of a . The source of the differences lies in not really knowing the exact equation of the function plotted and trying to eliminate the effects of absorption and

eccentricity with one and the same extrapolation curve. Although the errors vanish when $\theta = 90^\circ$, and therefore when $\cos^2 \theta$, $\cot \theta$ and $(\pi/2 - \theta) \cot \theta = 0$, the highest angle at which a line can be measured is set by the geometry of the camera and the wave-length of the radiation, and this angle is in the region of 80 to 85° . Because the extrapolation curves have different shapes, and the distances over which the extrapolations are carried out differ greatly, they cut the a -parameter axis in different places. To decide which form of the extrapolation curve we should take we must see how the line-contours and the eccentricity of the specimen influence the peak positions.

§2 ERRORS PRODUCED BY ECCENTRICITY AND ABSORPTION

(a) Errors produced by eccentricity

In (a) of figure 1 the specimen B is displaced from the centre of the camera A through the distance p at an angle ϕ with respect to the incident beam. Bradley and Jay (1932) have shown that the displacement may be considered as a vector split into the two components $p \sin \phi$ and $p \cos \phi$, as shown in (b) and (c) of figure 1.

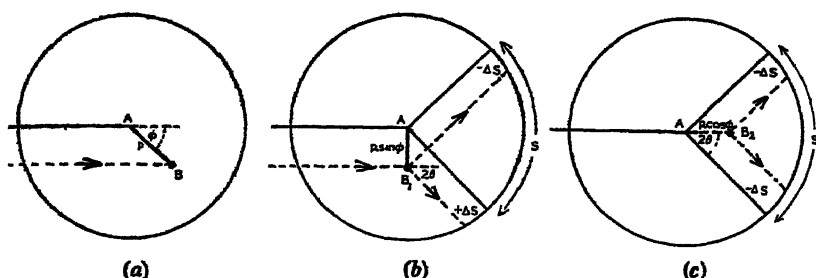


Figure 1. Effect of specimen eccentricity on line position.
(After A. J. Bradley and A. H. Jay, *Proc. Phys. Soc.* **44**, 563, 1932.)

The former displacement produces a change in S of $-\Delta S + \Delta S = 0$, while the latter produces a net change of $-\Delta S - \Delta S = -\Delta S_{\text{ecc}} = 2p \cos \phi \sin 2\theta$ in the value of S . Thus, in considering the errors introduced by the eccentricity of the specimen, we find that only displacements along the line of the undeviated incident beam have any effect.

(b) Errors produced by absorption

In a previous communication (1945) we showed how absorption displaced the peaks of the lines in the direction of greater θ by an amount which depended on the experimental conditions. If Δr is the shift of a diffraction line due to absorption alone, the error in S must be $\Delta S_a = 2\Delta r$, considering the two symmetrical portions of the film lying on each side of the incident ray. Hence the total line displacement due to absorption and eccentricity is

$$\begin{aligned} \Delta S &= \Delta S_a + \Delta S_{\text{ecc}} \\ &= 2\Delta r - 2p \cos \phi \sin 2\theta. \end{aligned} \quad (4)$$

The corresponding error in lattice parameter is, therefore,

$$\begin{aligned}\frac{da}{a} &= -\frac{dS}{4R} \cot \theta \\ &= -\frac{1}{4R} (2\Delta r - 2p \cos \phi \sin 2\theta) \cot \theta \\ &= -\frac{\Delta r}{2R} \cot \theta + \frac{\cos \phi}{R} \cos^2 \theta.\end{aligned}\quad (5)$$

Thus when θ tends to 90° , $\cot \theta$, $\cos^2 \theta$, and also Δr , all tend to zero, and, therefore, at $\theta = 90^\circ$, $da/a = 0$. To determine the most satisfactory form of extrapolation curve, we must obtain an expression for the variation of Δr with θ .

§ 3 FORMS OF EXTRAPOLATION CURVE

(a) The Bradley and Jay $\cos^2 \theta$ extrapolation

In figure 2 we show curves of Δr for different types of tube focus covering the range from $\theta = 0$ to 90° . They have been drawn for a Debye-Scherrer camera of radius $R = 95.0$ mm. and a distance AX of 150 mm. between the specimen

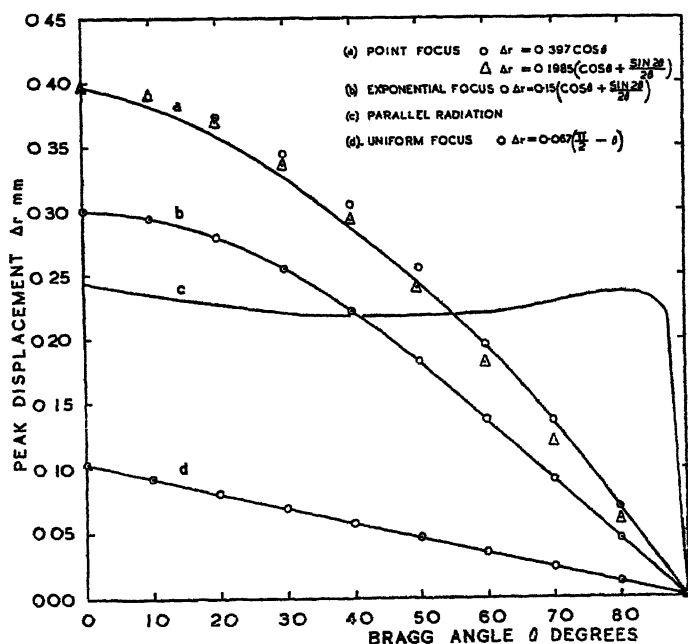


Figure 2. Values of Δr for different types of focus.

Full curves are those obtained from the line contours calculated for camera radius $R = 95.0$ mm., specimen-focus distance $AX = 150.0$ mm., specimen radius $r = 0.25$ mm., $\mu r = 2.0$. Points are values obtained by empirical formulae

and focus. The radius of the specimen, r , was taken to be 0.25 mm., a figure quite close to the one occurring in practice. The value $\mu r = 2.0$ was taken, the line-contours for this case being the ones most accurately determined. It should be borne in mind that Δr does not vary a great deal between the values $\mu r = 1.0$

and $\mu r = \infty$ although, of course, the macro-absorption factor α varies quite considerably.

If we compute the values of $\frac{\Delta r}{2R} \cot \theta$ for the cases of divergent radiation from uniform, exponential and point foci, we find they plot as near-linear functions of $\cos^2 \theta$ over the limited range from 60° to 90° . The R.H.S. of equation (5) is thus a near-linear function of $\cos^2 \theta$. By plotting a against $\cos^2 \theta$, it is possible to obtain almost straight extrapolation curves which simultaneously eliminate eccentricity and absorption errors. Provided we have sufficient spectra in the high-angle region, the $\cos^2 \theta$ extrapolation would seem to be the best one to use.

The possible case of parallel radiation is worthy of mention as it may arise if a perfect crystal is used as a means of producing monochromatic radiation.

The term $\frac{\Delta r}{2R} \cot \theta$ is then no longer a linear function of $\cos^2 \theta$, because of the abrupt manner in which Δr falls to zero between 85° and 90° . In such an instance the Bradley and Jay extrapolation against $\cos^2 \theta$ will no longer apply, and it then becomes necessary to correct all individual values of S by the amount Δr before computing the values of θ . The residual error in a would then be entirely due to the eccentricity factor alone and would simply be $\frac{da}{a} = -\frac{p \cos \phi}{R} \cos^2 \theta$, which plots as a straight line against $\cos^2 \theta$.

For comparison purposes it is easiest to plot curves of da/a rather than a against $\cos^2 \theta$. These are shown in figure 3 for conditions of parallel radiation and divergent radiation. The values for the eccentricity in each example are $p \cos \phi = 0$ and ± 0.5 mm.

For parallel radiation, the curves show a sudden upward inflexion below $\cos^2 \theta = 0.025$, which corresponds to the sudden change in Δr in the range $85^\circ < \theta < 90^\circ$. For divergent radiation, the extrapolation curves are very slightly convex upwards and are almost straight for the case of a uniform focus. This is no longer true if the plots are made against θ or $\cot \theta$ unless $p \cos \phi = 0$.

In their original derivation of the $\cos^2 \theta$ extrapolation rule, Bradley and Jay take an extreme case by assuming $\mu r = \infty$ and a point focus. The positions of the peaks are assumed to be near the high-angle outer edge of the lines at a fractional distance

$$\frac{\Delta r}{r} = \frac{\sin 2\theta}{2\theta} \left(1 + \frac{R}{AX} \right) \quad \dots\dots(6)$$

from the central ray. No formal proof is given for this relation, its sole justification apparently lying in the fact that when added to the eccentricity correction one arrives at the expression

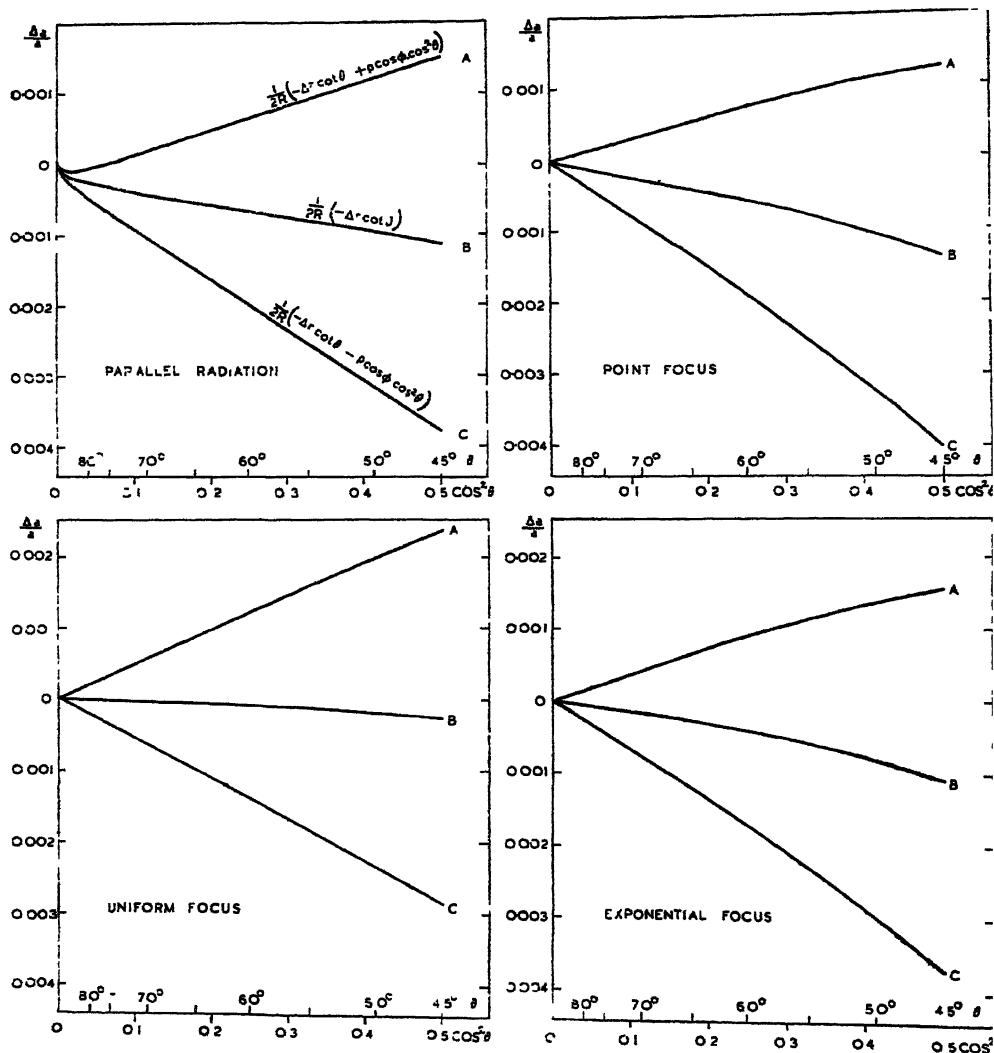
$$\frac{da}{a} = \left(\frac{p \cos \phi}{R} - \frac{r}{2\theta R} - \frac{r}{2\theta AX} \right) \cos^2 \theta, \quad \dots\dots(7)$$

which plots almost linearly against $\cos^2 \theta$.

Equation (6) gives results for $\Delta r/r$ rather different from our own, which are obtained directly from the line-contours. In the following table we make a comparison of $\Delta r/r$ calculated from equation (6) and the line-contour method for a point focus when $R = 95$ mm., $AX = 150$ mm. and $\mu r = \infty$.

Comparison of $\Delta r/r$ with $\mu r = \infty$ and point focus

θ ($^{\circ}$)	0	22½	45	50	60	70	80	90
Bradley and Jay	1.635	1.432	1.042	0.930	0.670	0.430	0.225	0.000
Contour method	1.635	1.510	1.185	1.115	0.910	0.665	0.370	0.000


 Figure 3. Plots of da/a against $\cos^2 \theta$.

$$(A) \quad \frac{da}{a} = \frac{1}{2R} (-\Delta r \cot \theta + p \cos \phi \cos^2 \theta),$$

$$(B) \quad \frac{da}{a} = \frac{1}{2R} (-\Delta r \cot \theta),$$

$$(C) \quad \frac{da}{a} = \frac{1}{2R} (-\Delta r \cot \theta - p \cos \phi \cos^2 \theta),$$

 with $AX=150.0$ mm., $R=95.0$ mm., $r=0.25$ mm., $\mu r=2.0$, $p \cos \phi=0$ and ± 0.5 mm

Although Bradley and Jay's results differ appreciably from our own, they follow substantially the same course, and it is for this reason that their approximation works. Their theory does not include the effects of a finite focal spot, which, as we have seen, has a major influence on the final positions of the peaks. Our investigation has shown that in the real case of a finite focus, the extrapolations against $\cos^2 \theta$ actually plot straighter than those given by the original theory for all values of $\mu r > 1.0$.

(b) *Other extrapolation possibilities—uniform focus, no eccentricity*

We have seen how over the limited range extending from $\theta = 60$ to 90° , the term $\Delta r / 2R \cot \theta$ of equation (5) was a near-linear function of $\cos^2 \theta$. This leads us to examine the curves of Δr against θ to see if we can express them as simple trigonometrical functions.

Consider first of all curve (d) in fig. 2, which refers to a uniform focus. This is a straight line, and its equation is of the form

$$\Delta r = k (\pi/2 - \theta). \quad \dots (8)$$

Thus in the absence of eccentricity, equation (5) reduces to the very simple form

$$\frac{da}{a} = -\frac{k}{2R} (\pi/2 - \theta) \cot \theta. \quad \dots (9)$$

If, then, the focal spot were uniform, as would most likely be the case in a gas tube, and if the camera were so accurately constructed that the specimen holder were at its geometrical centre, then the correct extrapolation procedure to use would be to plot a against the function $(\pi/2 - \theta) \cot \theta$. This would be a *perfectly straight-line plot* from 0 to 90° , and two reflexions, one in the region of 80° and one in the low orders, say 10 to 20° , would be quite sufficient to yield a high enough accuracy in the parameter determination. This opens up entirely new possibilities in the accurate determination of the lattice parameters of non-cubic crystals when very few suitable reflexions are available.

Exponential focus

The exponential focus with an intensity distribution of the form e^{-1-x^2} is probably the most prevalent type. The nature of the Δr curve given by such a focus is illustrated by (b) in figure 2. This curve can be matched *exactly* by the expression

$$r = k \left(\cos \theta + \frac{\sin \theta}{2\theta} \right), \quad \dots (10)$$

where $k = 0.15$ for the experimental conditions considered.

In the absence of eccentricity we have

$$\begin{aligned} \frac{da}{a} &= -\frac{\Delta r}{2R} \cot \theta \\ &= -\frac{k}{2R} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right). \end{aligned} \quad \dots (11)$$

In this particular case of exponential focus and no eccentricity, we obtain a perfectly linear extrapolation curve between $\theta = 0^\circ$ and $\theta = 90^\circ$ if we plot a against corresponding values of $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$. This plot should be even more valuable than the previous example given in equation (9), since an exponential focus is much more likely to be encountered in practice.

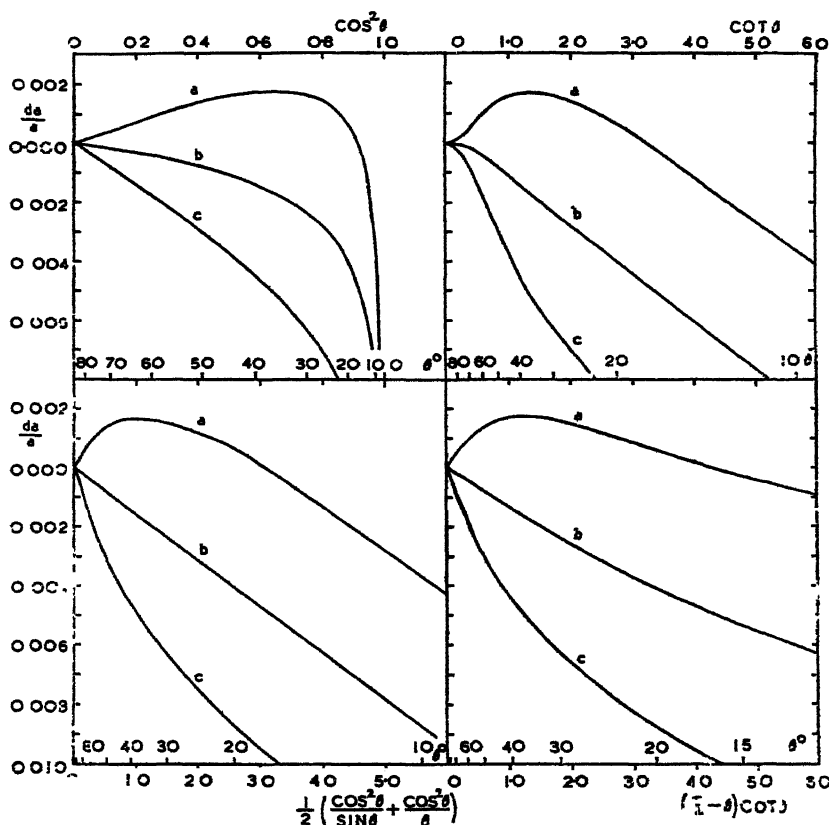


Figure 4 Typical extrapolation curves for exponential focus.

Camera radius $R = 95.0$ mm

Specimen-focus distance $AX = 150.0$ mm

Specimen radius $r = 0.25$ mm, $\mu r = 2.0$

(a) $p \cos \phi = 0.50$ mm.

(b) No eccentricity, or $p \cos \phi = 0$ mm.

(c) $p \cos \phi = -0.50$ mm.

When eccentricity is present, the error curve takes the form

$$\frac{da}{a} = -\frac{k}{2R} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right) + 2p \cos \phi \cos^2 \theta. \quad \dots (12)$$

This departs quite appreciably from the curve of equation (11) if plotted against $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ as shown in figure 4 (drawn for $p \cos \phi = \pm 0.5$ mm.). The

curvature is rather exaggerated by the crowding of the high-order reflexions into a region very close to the axis of ordinates.

If we increase the horizontal scale until it is comparable with that for the $\cos^2\theta$ plot, we find that in the range 30 to 90° the curves are not very different from each other. The great advantage to be gained by plotting a against $\frac{1}{2} \left(\frac{\cos^2\theta}{\sin\theta} + \frac{\cos^2\theta}{\theta} \right)$ is that in the absence of eccentricity the curve becomes a perfectly straight line and the fullest use can be made of the lowest orders.

It should not be difficult to make cameras free from eccentricity. Should eccentricity be present in an existing camera, it could be allowed for in the first stages of the θ calculations by correcting all values of S by the amount $\Delta S_{\text{ecc}} = -2p \cos\phi \sin 2\theta$. The magnitude $2p \cos\phi$ is a constant of the camera. It could be obtained by direct measurement or by plotting extrapolation curves for a cubic crystal and finding by trial a value of $2p \cos\phi$, which leads to a straight-line plot of a against $\frac{1}{2} \left(\frac{\cos^2\theta}{\sin\theta} + \frac{\cos^2\theta}{\theta} \right)$.

Point focus

This is an idealized case. The Δr curve illustrated by (a) in figure 2 can be matched over the range 60 to 90° by the expression

$$\Delta r = k \cos\theta. \quad \dots\dots (13)$$

Thus, in the absence of eccentricity,

$$\begin{aligned} \frac{da}{a} &= - \frac{k \cos\theta}{2R} \cdot \cot\theta \\ &= - \frac{k}{2R} \frac{\cos^2\theta}{\sin\theta} \end{aligned} \quad \dots\dots(14)$$

and a plot of parameter against $\frac{\cos^2\theta}{\sin\theta}$ is perfectly linear over the range 60 to 90°.

Since $\sin\theta$ changes very slowly, from 0.8660 to 1.0000 over this range, there is only a slight curvature in a $\cos^2\theta$ plot. This is, of course, the reason why Bradley and Jay's extrapolation proves to be so good.

An attempt was made to fit an expression of the form $\Delta r = k \left(\cos\theta + \frac{\sin 2\theta}{2\theta} \right)$ to the Δr curve for the point focus. The calculated points shown by triangles in figure 2 are rather high in the low orders and low in the high orders, with coincidences at 0°, 45° and 90°. The agreement is sufficiently close, even in this case, to justify an extrapolation of a against $\frac{1}{2} \left(\frac{\cos^2\theta}{\sin\theta} + \frac{\cos^2\theta}{\theta} \right)$.

§4 CONCLUSIONS

With careful camera construction there should be no eccentricity of the specimen. In that event perfectly linear extrapolation-curves can be drawn covering the range $\theta = 0^\circ$ to $\theta = 90^\circ$ for the real cases of uniform and exponential foci. Which type of plot should be used can easily be ascertained by taking a

powder photograph of a cubic crystal, plotting extrapolation curves of a against $(\pi/2 - \theta) \cot \theta$ and $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ and finding which gives a straight-line plot over the whole measurable range.

These straight-line extrapolation plots require fewer reflexions to obtain the same accuracy in spacing as given by $\cos^2 \theta$ curves, for the maximum use can be made of those low-order reflexions which lie in the region of 10° . Also, any "scatter" of the calculated parameters which personal errors introduce into film measurement can easily be allowed for by using the method of least squares to derive the most probable straight-line extrapolation curve.

In the past it has proved very difficult to make accurate parameter measurements on non-cubic crystals owing to the small number of reflexions with suitable indices. Instead, the complicated analytical method of M. U. Cohen (1935 and 1936) had to be employed. The straight-line plots described above should remove these difficulties.

It is felt that the new types of linear extrapolation-curve will enable lattice parameters to be measured with a much higher degree of accuracy than the one part in 50,000 which has hitherto been possible. It is quite probable that in the near future a greater precision in the determination of the x-ray wave-lengths will be required if full use is to be made of the increased accuracy in spacing determination. It will also be more necessary than ever to keep the temperature of the specimen constant while the powder photograph is being taken.

ACKNOWLEDGMENTS

The authors wish to thank the English Electric Company for their permission to publish this paper, and Dr. D. P. Riley, of the Cavendish Laboratory, Cambridge, for his very helpful discussions.

REFERENCES

- BRADLEY, A. J and JAY, A. H., 1932 *Proc. Phys. Soc.* **44**, 563 ; 1933 *Ibid.* **45**, 507.
 BUEGER, M. J., 1942 *X-ray Crystallography* (J. Wiley and Sons).
 COHEN, M. U., 1935 *Rev. Sci. Instrum.* **6**, 68 ; 1936. *Z. Kristallogr.* (A), **94**, 288.
 COMPTON, A. H. and ALLISON, S. K., 1935. *X-rays in Theory and Experiment* (London : Macmillan and Co., Ltd., 2nd edition), pp. 279 and 672.
 JETTE, E. R. and FOOTE, F., 1935. *J. Chem. Phys.* **3**, 10.
 KETTMANN, G., 1929. *Z. Phys.* **53**, 198.
 TAYLOR, A. and SINCLAIR, H., 1945. *Proc. Phys. Soc.* **57**, 108.
 WEGLE, J., 1934. *Helv. phys. Acta*, **7**, 46.
 WILSON, A. J. C. and LIPSON, H., 1941 *Proc. Phys. Soc.* **53**, 245.

RECENT IMPROVEMENTS IN A PRECISION BALANCE AND THE EFFICACY OF RHODIUM PLATING FOR STANDARD WEIGHTS

By JOHN JOB MANLEY,

Bournemouth; formerly Fellow of Magdalen College, Oxford

MS received 7 November 1944

ABSTRACT The main theme of this paper is a description of the improvements introduced into our precision balance with the object of enhancing its already high sensitivity and thus ensure a corresponding increase in accuracy in weighing.

As a secondary objective we have investigated the claim made for highly polished rhodium-plated weights and have found that weights so protected are superior to all others we have critically examined

§ 1. INTRODUCTORY

IN former papers* dealing with high-grade balances we have shown how certain of their inherent defects may be allowed for and a truer value obtained for any mass that is being weighed. In this present paper is given an account of more recent refinements, the introduction of which have enabled us to attain a degree of accuracy for a long time elusive and but recently realized. It may be noted that since the positions of the riders are each read to $1/34$ of a minor division of the beam, the smallest measurable difference in weight is $(0.01/34 \text{ mg.})$. We now describe in detail (1) the several refinements introduced, and (2) how these enabled us to apply exacting tests as to the advantage following the use of rhodium-plated weights.

§ 2. IMPROVEMENTS IN THE BALANCE

Our most recent improvements are:

- (α) the substitution of a new scale for that engraved upon the beam;
- (β) the use of pladuram wires for carrying the riders;
- (γ) the employment of platinum wire grids for ascertaining when the loaded pans possess a common temperature;
- (δ) a screened bridge wire; and, finally,
- (ϵ) a new cell for enclosing the pans.

We describe the several changes in the order just given

(α) *New scale for the beam.* For ensuring the highest order of accuracy in former researches it was necessary to calibrate the scale engraved upon the beam and apply corrections for various errors. To obviate this procedure, another

* *Philos. Trans. A*, 210, 387–415 (1910); *ibid.* 212, 227–260 (1912); *Proc. Phys. Soc.* 39, pt. 5, 444–448 (1927); *Proc. Roy. Soc. A*, 86, 591 (1912); "Balance", in Thorpe's *Dict. of Applied Chem.*, new edition (1937).

scale free from measurable discrepancies was obtained as follows:—First a thin strip of aluminum having the appropriate width and length was prepared and three holes drilled in it; one of these was at the centre and one near each end; that at the centre just admitted the passage of a screw, whilst the other two were slightly elongated horizontally; this done, the required scale was engraved with the aid of a dividing engine adjustable to 0.005 mm. and a lightly applied tool; the strip was then given a final polish and the scale calibrated with a travelling microscope. The result showed that within the limits of our measurements (0.005 mm.) the divisions were strictly equivalent. The scale was now secured to the beam by the central screw and the end screws driven so nearly home that although the strip made contact with the beam throughout, it was yet free to expand and contract consequent upon variations in temperature. It will however be granted that the slight difference in the expansions of aluminium and the phosphor-bronze composing the beam is wholly negligible.

(β) *The carrier wires for the riders.* Hitherto the riders designated A and B , the respective values of which are 10 and 1 mg., have been carried by platinum wires having a diameter of 0.1 mm. For these we have now substituted others of pladuram wire 0.07 mm. in diameter. It was found that the breaking load for this wire just exceeded 1 kg.; its tenacity was therefore 263, a value twice as great as that assigned to steel. The two wires were threaded through their respective riders, given the required, but not excessive, degree of tautness, and then, by means of nuts, permanently secured to the beam.

(γ) *New pan arrestors.* Before dealing with our next innovation, we remark that for ensuring high precision in weighing, a close approach to equality in the temperature of any two masses under comparison must obtain. In general this demand is met by placing the balance in a room for which variations in temperature are both small and slow; but for precision of the highest order stringent measures are imperative; and these must be such that we pass from assumption to certainty. The procedure was as follows. To begin with, the arrestors in use until now were in their entirety discarded for others constructed as shown in figure 1. Here the head h consists of a shallow brass cell having a copper cover c accurately fitting a recess. Two short quartz tubes, t_1, t_2 , pass through apertures in the floor of the cell and are secured therein with Faraday cement. The head of the screw s is supported by the quartz tube q_1 , to which it is cemented. Surrounding q_1 is a second tube q_2 , also of quartz, the lower end of which was opened by grinding so as to admit the stem of the inverted and glass-hard steel stud g , the curved surface of which was polished and which, when *in situ*, rests upon its operative cam. The concentric positions of q_1, q_2 and g were rendered permanent by a tightly fitting ring of asbestos at a , and by the application of cement round and about m , as indicated by the dotted lines. The arrestors were now ready for the reception of the prepared grids, each consisting of a spiral of platinum wire wound upon a mica former p . The diameter of the wire is 0.03 mm., and at 14° c. each grid has a resistance of 49.22 Ω . The grids were united to their respective and equal silver wire leads by means of short intermediary pieces of gold wire, the several junctions being effected by fusion. Each grid was now insulated by inserting it between two thin mica discs, then placed within its cell and enclosed by means of the copper cover c . Next the

arrestors were given their normal positions beneath the balance pans with their contained grids forming two arms of a Carey Foster bridge. The auxiliary coils of the bridge are silk-covered constantan wires; these are wound bifilarly upon a shellac-covered aluminium rod 1 cm. in diameter and protected with silk. This done, the supporting aluminium rod was mounted upon ebonite and fixed centrally behind the pillar of the balance. Initial measurements were now made with the aid of a Moll galvanometer, an instrument which was ultimately displaced by a 'short-period' galvanometer supplied by the Cambridge Instrument Co. The sensitivity of the new instrument proved to be considerably greater than that of the Moll, and on passing a current of 0.002 A. into the bridge, differences in the balancing point corresponding to 0.00002 Ω were measurable with certainty. Now a difference of 1° C in the temperature

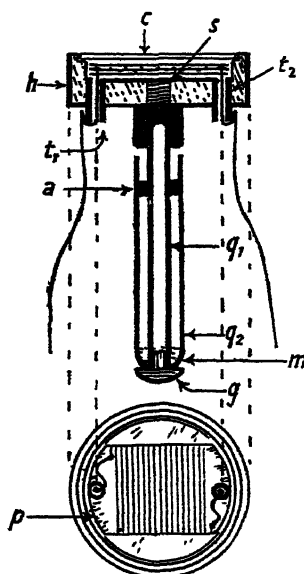


Figure 1.

in the two grids produces a difference in the resistances = 0.19 Ω ; hence variations of 1/10 000° C. can be detected

(8) *The encased slide wire.* This was designed to ensure uniformity in the temperature of the bridge wire and its slider. During preliminary tests it was found that even when the pans of the balance were contained within an aluminium cell common to both (*vide infra*), they generally exhibited small differences in temperature. These were ultimately found to be due to differences in the temperature of the 1 M. bridge. The difficulty was overcome by the use of a platinum-iridium wire 13 cm. long and 1.5 mm. in diameter, mounted and protected as shown in figures 2 and 3.

In the first figure, this wire is seen to be an air-line bent in the form of an arc, and the slider a radial arm ending in a platinum contact; the contact is maintained in action by a weak terminal spring. The binding screws connected with the leads, which pass through quartz tubes, are mounted upon ebonite fixed

to the aluminium base, whilst the base itself is secured to a vertical aluminium plate screwed to the lower surface of the galvanometer shelf. The completed apparatus is shown in figure 3. Here the enclosing aluminium shield, having walls 2 mm. thick, is seen to carry a circular scale some 5 cm. long and placed parallel to the wire within. The arm *a* is insulated by mica and ebonite, and across its circular window is fixed an attenuated fiducial wire, just visible in the figure.

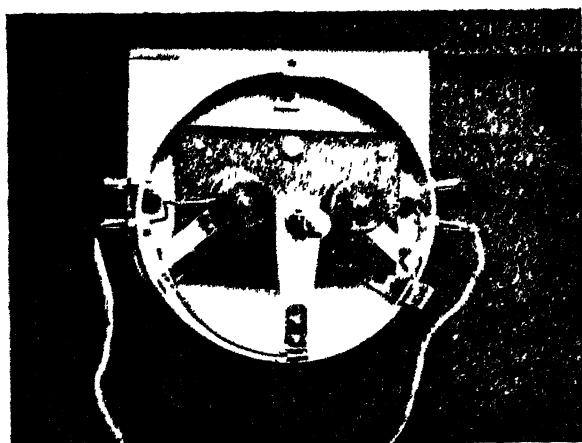


Figure 2.

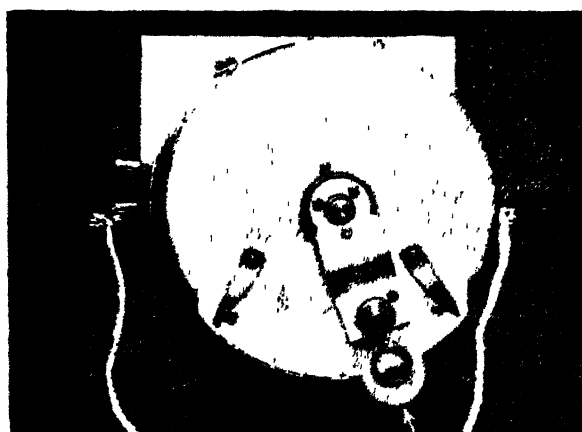


Figure 3

(*e*) *The cell enclosing the pans.* The new aluminium cell, which encloses both pans, so that their temperatures are equal, is depicted in figure 4. Its upper edge *AB* makes continuous contact with the lower surface of the aluminium base-plate of the beam chamber. Within a slot cut in the fixed plate *P* is shown a pointer *t*, a side view of which is given in *c*. This supplementary pointer is a thin aluminium wire ending in a knife-edge; it is given rigidity by passing it through a thin-walled and closely fitting capillary tube of quartz and finally attached horizontally to the normal pointer. Deflections are read by

means of a scale engraved as shown. Equilibrium of the loaded pans having been approximately established, the shutter *s* with its pivot at *o* is turned and the aperture closed, the relative values of the two masses are then precisely determined with the aid of a brightly illuminated scale and the vertical palladium mirror carried by the block of the central knife-edge. The pointer of light is 9 m. long and the sensitivity *S* of the balance equal to 461 mm. per 1 mg. The eye-piece of the telescope carries a micrometer with which deflections can be accurately read to 0.1 mm.; and if this be taken as the unit, $S=4610$. Again referring to figure 4, it will be seen that access to the pans is provided by the use of sliding shutters. One of these, *x*, is open, and the other *y* closed. Here it

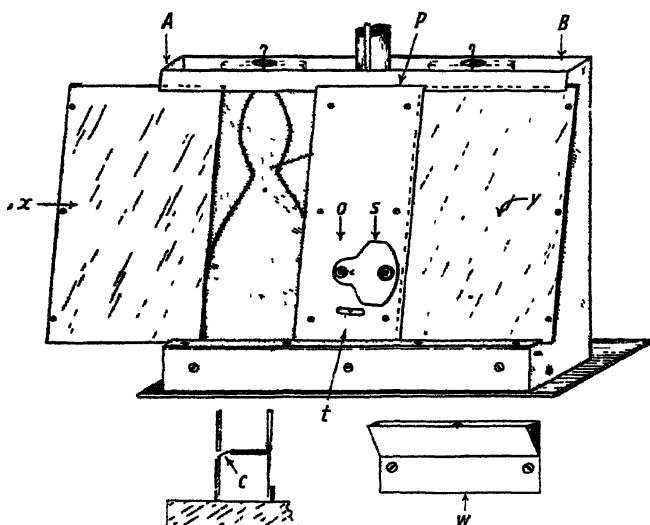


Figure 4

may be added that all interior surfaces are coated with vegetable black attached with the aid of copal varnish and permanently fixed by heating.

Lastly, immediately behind each pan suspension is a small weir, into which, from the posterior side, is delivered a slow stream of air purified by passing it successively through concentrated sulphuric acid, soda-lime and granulated calcium chloride *. One of these, *w*, is shown in figure 4. We conclude our descriptions of the most recent alterations in our balance by stating that the panels of the case, the top and also the base are completely enclosed with polished sheets of aluminium, and as a further aid to the establishment and maintenance of uniformity in temperature, the whole with the exception of the front sliding aluminium shutter, is surrounded with a wooden case having walls 2 cm thick.

§ 3 TESTS OF THE EFFECTIVENESS OF THE NEW IMPROVEMENT

Experiments designed to test the efficacy of the several innovations were now carried out. We begin by describing one made to ascertain the time required

* All new vulcanized rubber connections must be treated for the removal of volatile sulphur compounds. This is shown by the fact that by so doing 10 cm. of tubing having an internal diameter of 7 mm. yielded, when heated with a solution of pure sodium hydrate and hydrogen peroxide, 0.014 gm. of sulphur in the form of barium sulphate.

for a 100-gm. chromium-plated weight to cool after having been slightly warmed.

First, two similar 100-g. weights were placed in their respective pans and left overnight. On the following morning the ordinary air was displaced from the chamber (figure 4) enclosing the pans by streams of air purified as already described. Next, one of the two weights was held in the hand, and thus slightly warmed; it was then wiped with silk, replaced in its pan, the chamber closed and the streams of purified air maintained, this done, further procedure was as follows. Using a 1-m. Carey Foster bridge, in which the pan grids are the resistances under comparison, the slider was placed at a point corresponding to $0^{\circ} 7$ c. in excess of that of the second or unwarmed weight; then, when the chosen initial temperature had been nearly reached, the circuit was kept closed and the

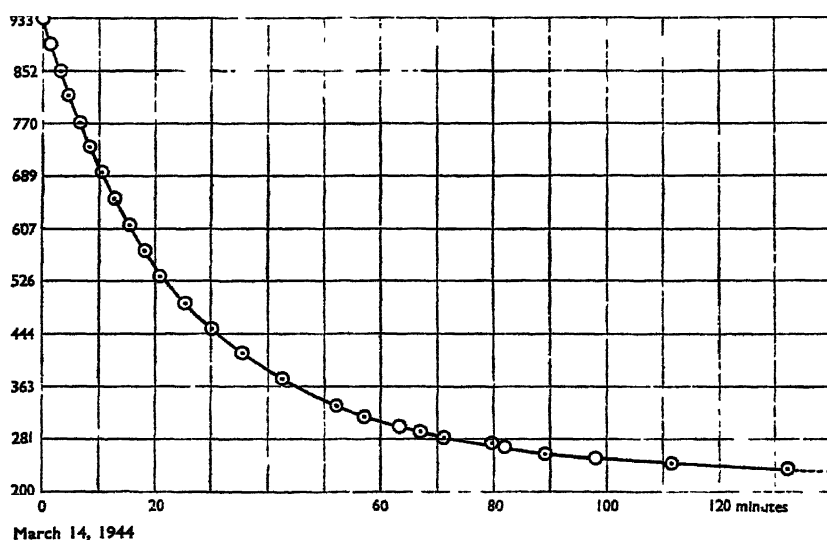


Figure 5 Cooling curve for 100-gm Cr-plated weight placed within chamber common to both pans. Data obtained after all final precautions had been completed. A second 100-gm Cr-plated weight not handled was in the other pan.

drift of the galvanometer deflection followed, and as this crossed zero the chronometer was started; then the slider was moved forward to a first predetermined point and the time noted for a second zero reading.

Proceeding thus some 26 zero points were obtained. From the data now acquired, the cooling-curve figure 5. was drawn. And here it may be noted (1) that the curve is almost perfectly smooth, and (2) that although after a period of 130 minutes, equilibrium of temperature had not been established, effects following the want of it are insignificant and are treated as non-existent. In this experiment the reassumption of a temperature common to both weights required a period approaching a maximum. For other weights having more highly radiating surfaces the time for cooling would be less.

We now deal with changes in the sensitivity of the balance for various loads.*

* Here it may not be inappropriate to emphasize again the merits of a phosphor-bronze cantilever beam having its terminal knife-edges *within* instead of *beyond* the outer struts.

Hitherto, in so far as we could discover, the sensitivity of the balance had been one and the same for all loads, but now that this had been largely increased, it was deemed necessary to retest the beam. Accordingly this was done, use being made of loads ranging in value from 0 to 200 g. The results are represented in figure 6.

The changes, small in themselves, and varying within the limits of 6×10^{-4} and 7×10^{-5} mg. must nevertheless be taken into account when seeking to ensure the highest order of accuracy. We remark that the steep part of the curve indicates an initial and very slight relative readjustment of one or more of the knife-edges, the probability being that the central one only was affected. We conclude by drawing attention to some experiments made to test the relative merits of the two methods used for screening the pans.

In one method we attempted to establish a common and uniform temperature in and about the pans and their contents by protecting them not only with their individual brass cylinders, but also by enclosing the sides, the front sliding panel,

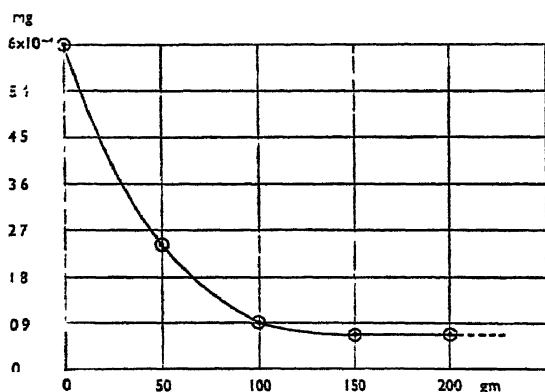


Figure 6 Sensitivity of balance.

and also the top of the case, with sheets of bright aluminium. In the other we relied upon a theoretical efficacy resulting from the use of a chamber (figure 4) common to both. For each group of experiments the pans were unloaded, this to ensure quick response.

The first series of experiments was carried out on a bright and sunny day, the second during an overcast sky and the third on a day when calm and rainy weather prevailed. The results are shown graphically in figure 7.

It will be seen that the several maximum variations in temperature of the two pans are respectively $0^{\circ}075$, $0^{\circ}030$ and $0^{\circ}012$ c. Now these variations in temperature would affect the apparent weight of a body having a volume of 100 c.c. to the extent of $\pm 30 \times 10^{-6}$, 12×10^{-6} and 4.8×10^{-6} respectively, and all these come within the range of the balance and can be measured with certainty. We now pass to a brief consideration of our final series of experiments made with the pans enclosed within one and the same chamber (figure 4), and for which we used the circular bridge wire (figures 2 and 3).

Again, the pans being empty, several series of observations were carried out on succeeding *sunny* days. Within that time the temperature varied by some

6° C.; yet, notwithstanding this, the mean variations in the temperature of the two grids and, therefore, of the pans was $\pm 0^{\circ}0004$ C only. This is 1/30th of the *minimum* difference noted during the corresponding experiments of the first series, and the effect is as will be seen, wholly negligible. A graphical

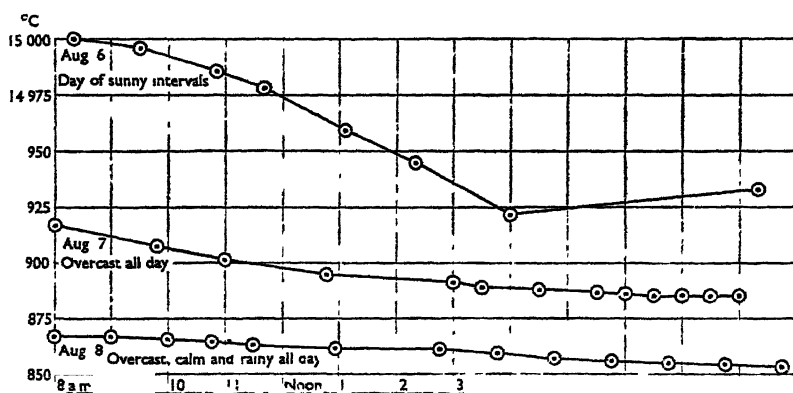


Figure 7.

representation of the data results, if put forth on the scale used in figure 7, in a right horizontal line. Having now achieved our first and main objective, we pass to the second and minor problem, namely, as to whether rhodium, used as a protective, surpasses chromium, hitherto found to excel all others.

§ 4 THE EFFICACY OF RHODIUM-PLATING

For this investigation there were available two sets of rhodium-plated weights, the one by Oertling and the other by Bunge. Each set consisted of weights ranging in value from 1 to 500 gm., and both were stored as described in a former paper.* Measurements were carried out with the heavier weights only, this, owing to their larger surfaces, offering a greater mass of the film to be determined. We were however, limited to the use of weights of 200 gm. the maximum load for which the balance had been built. Our inquiry was, however, somewhat enlarged by the examination of three 100-gm. weights. One of these was coated with rhodium and another with chromium; the third consisted of the alloy known as nichrome. All these weights were, as usual, first thoroughly wiped with silk which had been successively cleansed with dilute ammonia and distilled water, then dried and finally washed with highly purified benzene and again dried. The wiped weights were then brushed with newly cleansed camel's hair and placed within their respective cells. A period of 18 months having been allowed for storage, the several weights were successively tested for any film that might have been acquired. As usual, two equal weights (say 200 and 200), were lifted from their cells, lightly brushed with camel's hair and placed the one in the L-pan and the other in the R-pan of the balance; then, observing all known precautions, a number of comparisons were made during the two or three ensuing days. The several series of determinations before wiping having been completed, one of the two weights (say 200), was transferred

* *Phil. Mag.* 19, 243 (1935).

to the silk duster and gently rubbed with the same; this done, the weight was again brushed with camel's hair to remove all loose particles, then replaced in the pan whence it had been taken, and a second and final series of weighings carried out. Knowing the temperature coefficient of the balance for the given load, it was now possible to reduce all the observations for a common temperature, and this was accordingly done. with the results set forth in the following table:—

Weight	Maker	Coating	Period	Area of weight	Total film	Film per 1 cm. ²
			months	cm. ²	mg.	mg.
200	Oertling	Rhodium	18	53	·00212	·00004
200	"	"	18	53	±0	0
200	Bunge	"	18	46	±0	0
200	"	"	18	46	±0	0
*100	Oertling	"	69	34	·01292	·00038
† 100	"	Chromium	19	27	·04773	·00177
‡ 100	"	Nichrome weight	26	30·3	·00424	·00014

* The charcoal beneath the weight was not re-heated during storage.

† Stored in a cell of African blackwood in absence of charcoal.

‡ Stored over charcoal which was heated at intervals

From the above tabulated results we note that of the four major weights tested, the first alone was adversely affected, but even so, the observed loss approximated to 1 in 10×10^7 parts only.

Regarding the tests made with the three 100-gm. weights, we remark that these afforded an opportunity for ascertaining whether their values were affected under the several conditions imposed and, if so to what extent. In the first it will be seen that even when a weight is rhodium-plated it yet acquires a definite film unless the charcoal is periodically heated. In the present case the film was some three times greater than one that is normal.

The chromium-plated weight stored in African blackwood acquired a film some 90 times larger than that usually found. Here, then, we have an example illustrating the important rôle played by coconut charcoal.

In conclusion, we remark that the film found upon the nichrome 100-gm. weight, after it had been stored for 26 weeks in charcoal, was strictly normal and, therefore, calls for no further comment.

REVIEWS OF BOOKS

The Simple Calculation of Electrical Transients, by G. W. CARTER. Pp. viii + 120. (Cambridge : University Press, 1944.) 8s. 6d. net.

The sub-title, *An Elementary Treatment of Transient Problems in Linear Electric Circuits by Heaviside's Operational Methods*, together with the restriction (stated in the introduction) to circuits with "lumped" parameters, sufficiently indicates the scope of the book. The impedance and admittance operators are rational algebraic fractions which can be decomposed into partial fractions of a few standard types, so that rules for interpretation can be concisely and completely given. But a large number of interesting and important problems do lie within this restricted field, and each advance of technique is illustrated by the solution of a problem of real practical significance. The freshness and practicality of the worked examples is one outstanding merit of the book.

There is no claim to "make things easy", but any interested student with a knowledge of mathematics up to that of an average engineering graduate should be able easily to master the contents of this book. Indeed, it would seem to the present writer to be the most successful *elementary* treatment he has so far met. Its value would, however, be very considerably enhanced by the provision, say at the end of each chapter, of a few exercises for the student to work for himself.

A useful series of appendices contains, besides a list of interpretations of the simpler operational forms, formulae in trigonometry, calculus, and theory of equations relevant to the subject—the last including Routh's criteria of stability. W. G. B.

Five-figure Logarithm Tables. Pp. 1-73. i-xx. 30-119. (Published for the Ministry of Supply by His Majesty's Stationery Office, 1944.) 7s. 6d. net.

This compilation is, as explained in the preface, a war-time measure, and consists of reprints of existing five-decimal tables of logarithms of numbers and trigonometric functions.

The first portion (pp. 1-73) is a reprint, from the stereos of Chappell's table of logarithms, which gives five-decimal mantissae of the logarithms of numbers from 10000 to 40000 and from 4000 to 10000. Modal differences per line are given, and as these never exceed ten (in units of the fifth decimal), interpolation is linear, and might well be mental, although proportional parts are provided in the second portion of the table.

The third portion is a photographic reprint of pp. 30-119 of Bremiker's table of five-decimal logarithms of sines, tangents, cotangents and cosines for the range $0^{\circ}00$ ($0^{\circ}01$) $45^{\circ}00$ —i.e., every hundredth of a degree from 0° to 45° . It is taken from the sixteenth, stereotyped, edition, dated 1925. First differences are given, and proportional parts for angles greater than 3° .

The "difficult" range ($0^{\circ}5^{\circ}$ for sines and tangents, and $85^{\circ}90^{\circ}$ for cosines and cotangents) is covered by von Rohr's addition, paged i-xx, to Bremiker, consisting of logarithms of sines and tangents for the range $0^{\circ}000$ ($0^{\circ}001$) $5^{\circ}000$ —i.e., every thousandth of a degree from 0° to 5° . Differences are not given, but proportional parts appropriate to angles of upwards of $0^{\circ}5$ are.

Explanatory matter has been omitted, on the justifiable ground that potential users will be sufficiently experienced not to need it.

There is no point in criticizing the tabular material. It is by no means new, and the best available has been chosen.

The fact of greatest significance is the use of degrees and *decimals* of a degree in the trigonometrical tables. This practice has advantages, and seems to be rapidly superseding the use of minutes and seconds in many quarters. The need by the optical industry for a "fine-grained" table with such a subdivision of the degree seems to have been the primary incentive to the production of this volume. It is surprising that, if the need were great, the production has been delayed until after five years of war. Now that these tables are available, others also may find them useful.

It is a very great pity, however, that Bremiker's table of S and T , for converting $\log A^\circ$ into $\log \sin A^\circ$, or $\log \tan A^\circ$, p. 170 of the original, and referred to in the P.P. column on p. 30, was not also reproduced. Linear interpolation ceases to be adequate in von Rohr's table below about $0^\circ 2'$, this table does not completely solve the interpolation problem, but the S and T tables do. If this one-page table were reproduced separately, say on thin card, it would enhance the usefulness of the book. A separate card would, indeed, be more convenient in use than a page bound up with the volume. W. G. B.

What is Life ?. by ERWIN SCHRODINGER. Pp. viii + 91. (Cambridge: University Press, 1944.) 6s. net.

The middle section of this little book gives a clear and, to a physicist, convincing account of genetical principles and of the Delbruck model of a gene. According to this model, a gene is a very large single molecule, and a mutation, spontaneous or induced by x-ray dosage, is a quantum transition of the molecule to a new isomeric form. From his discussion of the Delbruck model, the author draws the general conclusion "that living matter, while not eluding the 'laws of physics' as established up to date, is likely to involve 'other laws of physics' hitherto unknown, which, however, once they have been revealed, will form just as integral a part of this science as the former". The final chapters are concerned with clarifying this conclusion, which, we are told, provided the only motive for writing the book. While these chapters and the opening chapter enlarge, in an interesting way, on the statistical element in physical laws and on the "statistical tendency of matter to go over into disorder" as contrasted with the organism's astonishing gift of concentrating a 'stream of order' on itself, they leave the main conclusion still rather obscure. The inference from the Delbruck model is presumably that, when the principles of quantum theory are included among the laws of physics, the behaviour of living organisms becomes explicable. What place then remains for "other laws of physics" unless the title of "laws of physics" is conferred on the regularities of biological behaviour which are explained?

The reconciliation of Determinism and Free Will is considered in a five-page Epilogue, in which the Upanishads, double personality, Schopenhauer, plurality of souls, Kant and the tree in the quad, all have their reference—a fairly rich philosophical bolus.

In the difficult field between physics and biology a clear-cut argument is not to be required, and although the reader, like the reviewer, may find the author's main point (quoted above) somewhat too subtle and not unconnected with an ambiguous use of the term "physical law", he can hardly fail to find the book, as a whole, illuminating and stimulating. W. S. S.

CORRIGENDUM

In *Proc. Phys. Soc.* 57, part 1, the first line of p. 19 should follow the first line of p. 18 and precede the line appearing as the second of p. 18.

RECENT REPORTS AND CATALOGUES

Co-operative Electrical Research. Pp. 62. 1944. THE BRITISH ELECTRICAL AND ALLIED INDUSTRIES RESEARCH ASSOCIATION, 15 Savoy Street, London W.C.2.

Report on the Needs of Research in Fundamental Science after the War. Pp. 61. Printed for private circulation, January 1945. THE ROYAL SOCIETY, Burlington House, Piccadilly, London W.1.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 57, PART 3

1 May 1945

No. 321

THE THEORY OF INDENTATION AND HARDNESS TESTS

By R. F. BISHOP, R. HILL

AND

N. F. MOTT, F.R.S.

Received 2 January 1945, read 9 March 1945

ABSTRACT A discussion is given of the indentation of ductile materials by cylindrical punches with conical heads. On the experimental side, experiments have been made with work-hardened and with annealed copper, with penetrations up to nine times the diameter of the punch. It is found that the load rises towards a maximum value which is not approached until the base of the cone has travelled four to five diameters into the copper block. Denoting this maximum load by $p_0 A$, where A is the area of the cross-section of the punch, it is found that p_0 for a lubricated punch is about twice the hardness, or five times the yield stress, of the work-hardened material. A theoretical method is given for calculating p_0 , as follows: the pressures p_c and p_s required to enlarge a cylindrical and a spherical hole in a material showing any kind of strain hardening can be calculated. It is plausible to assume that p_0 should be between p_c and p_s , and since p_s is only slightly greater than p_c , an approximate theoretical estimate of p_0 is obtained. This is in good agreement with experiment. In the light of these results a qualitative discussion is given of hardness testing, and it is shown both on experimental and on theoretical grounds that with lubricated cones and work-hardened materials the hardness, i.e. load/indentation area, will not depend much on the angle of the cone unless this is less than 10° .

§ 1. INTRODUCTION

WHEN a punch, cone, or sphere or other indenting tool is forced into a ductile material, the load required to form a given indentation will depend on the following factors:

- (a) The shape and size of the indenting tool, and of the specimen if not sufficiently large.
- (b) The coefficient of friction between the indenting tool and the material.
- (c) The yield point and strain-hardening properties of the material, which will depend on temperature and rate of loading.

Up to the present it has proved impossible to develop the mathematical theory of plasticity to the point where it would be possible to calculate the load, and thus any of the conventional hardness numbers, in terms of these factors. In this paper, however, we shall show that it is possible to calculate an approximate value for the load required to force a cylindrical punch deep into a semi-infinite block of ductile material; by deep is meant a penetration equal to four or five times the diameter of the punch. We shall also describe some experiments on static punching into copper blocks designed to test the theory. In the light of these results a general discussion is given of hardness measurement.

§2. EXPERIMENTS ON DEEP PUNCHING*

For the material to be indented, blocks of copper, 3.75 in. in diameter and 1.4 in. long were used. The copper was generally in the strain-hardened condition, though one set of measurements was made with annealed copper.

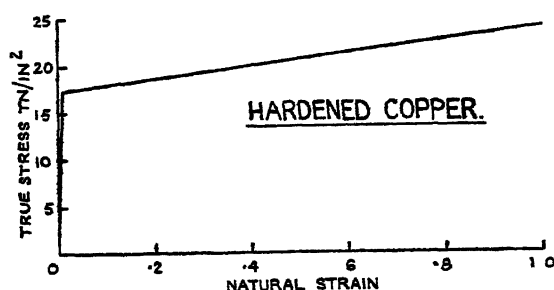


Figure 1.

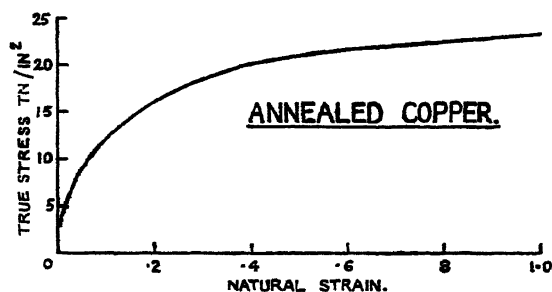


Figure 2.

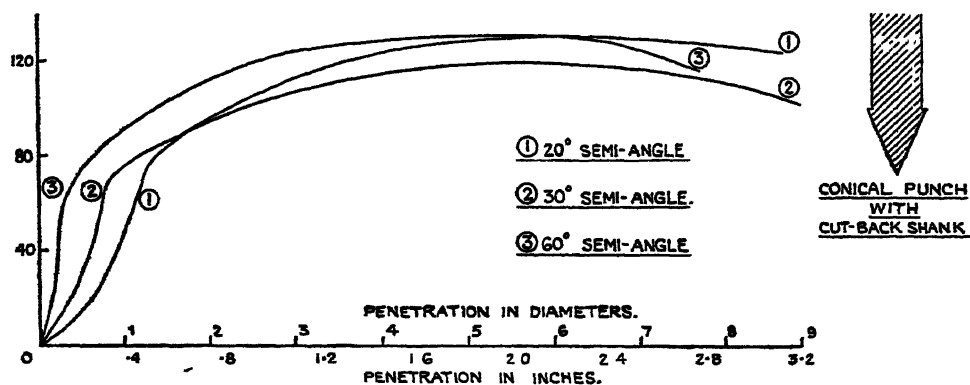


Figure 3. Unlubricated punching into hardened copper.

Stress-strain curves for the copper in compression are shown in figures 1 and 2.† The punches used were 0.350 in. in diameter and had conical heads with semi-angles of 20°, 30° and 60°. The diameters of the parallel parts of the punch

* These experiments were carried out at the Engineering Laboratory of the University of Cambridge under extra-mural contract to the Ministry of Supply.

† For definition of natural strain, cf. equation (1), p. 151.

were made 0.004 inch less than the head diameter; this prevented rubbing of the parallel portion of the punch against the sides of the hole and consequent increase of load due to friction. A diagram of the punch head is shown in figure 3. Punching was carried out at a rate of 1.5 in. per hour approximately.

In figure 3 we show, for the hardened copper, the measured load divided by the cross-sectional area of the punch ($\pi \times 0.175 \times 0.175$ sq. in.). This quantity has the dimensions of a pressure. Figure 4 shows the same quantity

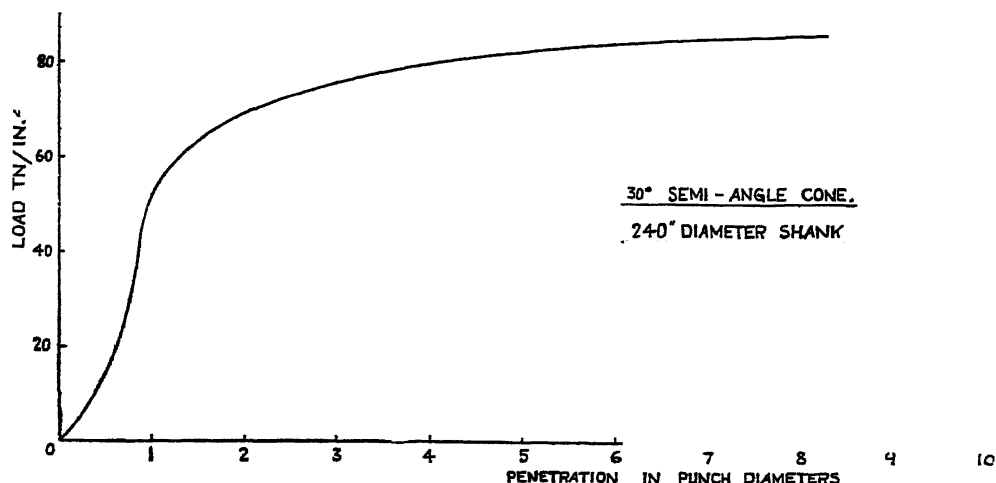


Figure 4 Unlubricated punching into annealed copper.

for annealed copper. It will be seen that as the penetration increases beyond three or four punch diameters, the load tends to a stationary value, which (for the hardened copper) is not markedly different for the three punches. There is some fall-off as the punch approaches the back of the specimen. It is this maximum load which we shall attempt to calculate in the next sections. It is less for annealed than for hardened copper.

§3. EFFECTS OF FRICTION

Various attempts to lubricate the surface, and thus to measure the coefficient of friction, were made; the most successful was by rotating the punch at a speed of 100 revolutions an hour. The frictional force would then have acted approximately at right angles to the direction of the punch. Curves showing load against penetration for rotated punches into hardened copper are shown in figure 5. That for the 20° punch shows the greatest tendency to approach a constant load; it was suspected that for the other punches the rotation became unstable and that the shank rubbed against the surface of the hole. The penetrations were continued until the punches fractured.

The difference between the ultimate resistance to penetration with rotated and unrotated conical-headed punches of 20° semi-angle appears from the graphs to be about 45 tons/sq. inch.

It is reasonable to assume that for the high pressures concerned the frictional force s per unit area is independent of pressure, and of the order of the shear stress of the material. The shear stress is $\frac{1}{2}Y$, where Y is the yield stress and

equal to 17.5 tons/sq. in. for the hardened copper used in these experiments. For annealed material s , however, should be rather greater than $\frac{1}{2}Y$, because the copper at the surface will be subject to very considerable cold work.

The frictional force due to the conical portion of a punch of semi-angle θ and cross-sectional area A will then be $As \cot \theta$.

Writing $s \cot 20^\circ = 45$ tons/sq. inch, we find $s = 16.5$ tons/sq. inch.

This value, if correct, leads to the following pressures for the other cones, obtained by subtracting $s \cot \theta$ from the values in figure 3:

$$\theta = 20^\circ : 130 - s \cot 20^\circ = 85 \text{ tons/sq. inch.}$$

$$\theta = 30^\circ : 119 - s \cot 30^\circ = 91.5 \quad ,,$$

$$\theta = 60^\circ : 131 - s \cot 60^\circ = 121 \quad ,,$$

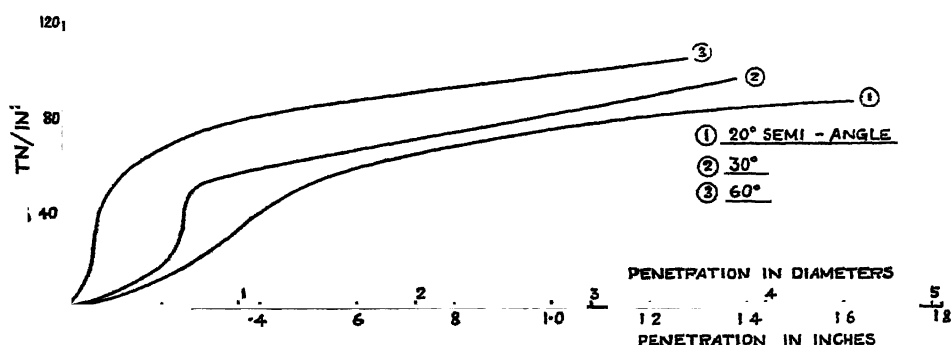


Figure 5 Punching with rotation into hardened copper

The value for $\theta = 60^\circ$ is probably not significant: it is known that a blunt punch pushes in front of itself a plug which acts effectively as the indenting tool, so that the simple frictional correction described here is not valid. The difference in pressure between the 20° and 30° cones is hardly outside the limits of experimental error.

Since the frictional force is large, the approximation used here of considering it as without effect on the stress component perpendicular to the surface of contact may lead to some error, which, however, we are unable to estimate.

§ 4. THEORETICAL ESTIMATE OF THE LIMITING LOAD

In the last section we have seen how experiments show that the load on a lubricated punch appears to rise to a limiting value, which, for hardened copper, is equal to about five times the yield stress multiplied by the area of the punch; also that this value is not approached until after a penetration of several diameters. In this section we attempt a theoretical estimate of this limiting load.

A full theoretical solution of the equations determining the strains round the head of the punch has not proved possible. Solutions of the following problems can, however, be found:

- (a) Starting with a small hollow sphere in the body of an infinite block of ductile material capable of work-hardening, to determine the pressure p_s which, when applied to the surface of the hole, will enlarge it indefinitely by plastic flow of the material.

- (b) Starting with a cylindrical hole of infinite length, to find the pressure p_c that will enlarge the hole indefinitely.

Calculation shows that $p_s > p_c$, but that the difference between them is only about 15 per cent of either. The pressure on a lubricated punch deep in a material is likely to lie between p_s and p_c ; for a very sharp punch (small θ) it should approach p_c , while for a blunt punch the material is pushed away from the head of the punch in much the same way as from the neighbourhood of an expanding sphere, so the pressure should be near to p_s . Thus the fact that p_s and p_c are nearly equal enables rather a close estimate of the limiting load to be made.

We shall now calculate the pressures p_s and p_c .

In both problems we assume that the material has a yield stress Y ; for stresses less than Y the material is assumed to be elastic, with Young's modulus E and Poisson's ratio ν . Above the yield stress we shall assume that in compression the true stress σ is given by

$$\sigma = Y + f(\epsilon), \quad \sigma \geq Y \quad \dots\dots (1)$$

where ϵ is the natural strain, given by

$$= \log_e \left(\frac{\text{final area of cross-section}}{\text{initial area of cross-section}} \right).$$

At any stage in the expansion of the spherical and cylindrical hole we denote the radius of the hole by a , and the radius of the plastic region surrounding it by c . Within the plastic region we shall neglect all volume changes.

We shall show below that

$$\begin{aligned} \frac{c}{a} &= \left\{ \frac{E}{(1+\nu)Y} \right\}^{\frac{1}{2}} && \text{Spherical hole} \\ &= \left\{ \frac{\sqrt{3}E}{2(1+\nu)Y} \right\}^{\frac{1}{2}} && \text{Cylindrical hole} \end{aligned} \quad \dots\dots (2)$$

Thus as the hole expands the ratio of the radius of the plastic region to that of the hole remains constant.

In order to progress further with our problem it is necessary to make some assumption about the stress-strain relation in the plastic region for a more general case than simple tension or compression. Following Nadai (1937), we shall assume that the octahedral stress is a definite function of the octahedral shear strain for all strain configurations; this is at any rate likely to be the case when, as here, the principal axes of stress and strain in a given element coincide and do not rotate during plastic flow. The octahedral stress τ_n is defined by the equation

$$\tau_n = \frac{1}{3} \{ (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2 + (\sigma_1 - \sigma_2)^2 \}^{\frac{1}{2}}. \quad \dots\dots (3)$$

and the octahedral natural shear strain γ_n by the differential relation

$$d\gamma_n = \frac{2}{3} \{ (d\epsilon_2 - d\epsilon_3)^2 + (d\epsilon_3 - d\epsilon_1)^2 + (d\epsilon_1 - d\epsilon_2)^2 \}^{\frac{1}{2}}, \quad \dots\dots (4)$$

where $\sigma_1, \sigma_2, \sigma_3$ are the principal stresses and $\epsilon_1, \epsilon_2, \epsilon_3$ are the principal natural strains. In a tensile or compression test

$$\tau_n = \sqrt{2}\sigma/3, \quad \gamma_n = \sqrt{2}\epsilon;$$

the universal relation between τ_n and γ_n then follows from (1) and is

$$\frac{3}{\sqrt{2}} \cdot \tau_n = Y + f \left(\frac{\gamma_n}{\sqrt{2}} \right), \quad \tau_n \geq \frac{\sqrt{2}Y}{3}. \quad \dots (5)$$

In our cylindrical problem, let σ_r , σ_θ be the radial and tangential stresses. Following Nadai, we set for σ_z in the plastic region

$$\sigma_z = \frac{1}{2}(\sigma_r + \sigma_\theta).$$

This clearly follows from the assumptions of plane strain and no volume change. We thus obtain from (3)

$$\tau_n = \frac{1}{\sqrt{3}} \sqrt{\frac{3}{2}} \cdot (\sigma_\theta - \sigma_r).$$

Also, since the z component of strain is zero and $(\epsilon_\theta + \epsilon_r)$ vanishes by our assumption of no volume change,

$$\gamma_n = 2\sqrt{\frac{2}{3}} |\epsilon_r|.$$

It therefore follows from (5) that the stress-strain relation is

$$\sigma_\theta - \sigma_r = \frac{2Y}{\sqrt{3}} + \frac{2}{\sqrt{3}} f \left\{ \frac{2}{\sqrt{3}} |\epsilon_r| \right\}; \quad \sigma_\theta - \sigma_r \geq \frac{2Y}{\sqrt{3}}. \quad \dots (6)$$

In the boundary of the plastic region

$$\sigma_\theta - \sigma_r = \frac{2Y}{\sqrt{3}}. \quad \dots (7)$$

This is the usual Mises condition for plasticity.

If we integrate the condition of equilibrium, namely

$$\frac{d\sigma_r}{dr} = \frac{\sigma_\theta - \sigma_r}{r}, \quad \dots (8)$$

throughout the plastic region from $r=a$ to $r=c$, we obtain for the pressure p_c on the boundary of the hollow cylinder (the value of $(-\sigma_r)$ at that point)

$$p_c = \frac{Y}{\sqrt{3}} \left(1 + 2 \log \frac{c}{a} \right) + \frac{2}{\sqrt{3}} \int_a^c f \left\{ \frac{2}{\sqrt{3}} |\epsilon_r| \right\} \frac{dr}{r}. \quad \dots (9)$$

To evaluate the integral we need an expression for ϵ_r . If an element at radius r in the plastic region when the hole is of radius a was formerly at radius s when the hole was of zero radius, then

$$r^2 = a^2 + s^2$$

and

$$\epsilon_r = -\epsilon_\theta = -\log (r/s).$$

Thus (9) becomes

$$p_c = \frac{Y}{\sqrt{3}} \left(1 + 2 \log \frac{c}{a} \right) + \frac{2}{\sqrt{3}} \int_a^c f \left\{ \frac{2}{\sqrt{3}} \log \frac{r}{\sqrt{r^2 - a^2}} \right\} \frac{dr}{r}. \quad \dots (10)$$

The evaluation of the integral will be carried out in the next section.

We have next to find the value of c/a ; to do this we equate the plastic and elastic displacements (denoted by u) on the plastic-elastic boundary. The assumption of no volume change in the plastic region gives, for the displacement on the boundary of the plastic region,

$$u = a^2/2c.$$

On the boundary of the elastic region the assumption that $\sigma_\theta - \sigma_r = 2Y/\sqrt{3}$ gives

$$u = (1 + \nu)Yc/\sqrt{3}E.$$

Equating these gives

$$\frac{c^2}{a^2} = \frac{\sqrt{3}E}{2(1+\nu)Y}. \quad \dots\dots(11)$$

A similar analysis applies to the expansion of a spherical hole; it is found that

$$\tau_n = \sqrt{2}(\sigma_\theta - \sigma_r)/3,$$

and, since $\epsilon_r + 2\epsilon_\theta = 0$,

$$\gamma_n = 2\sqrt{2} \cdot \log(r/s).$$

It follows that

$$\sigma_\theta - \sigma_r = Y + f\{2 \log(r/s)\}.$$

The condition at the plastic-elastic boundary is now

$$\sigma_\theta - \sigma_r = Y,$$

and the equilibrium equation

$$\frac{d\sigma_r}{dr} = \frac{2(\sigma_\theta - \sigma_r)}{r}.$$

On integrating the latter equation we obtain

$$p_s = \frac{2Y}{3} \left(1 + 3 \log \frac{c}{a}\right) + 2 \int_a^c f \left(2 \log \frac{r}{(r^2 - a^2)^{\frac{1}{2}}}\right) \frac{dr}{r}, \quad \dots\dots(12)$$

where

$$\frac{c^2}{a^2} = \frac{E}{(1+\nu)Y}.$$

§ 5. EVALUATION OF INTEGRALS

The true-stress natural-strain curve for many materials becomes straight at large strains (*cf.* MacGregor and also figure 2). Figure 7 shows the tensile curve for mild steel (MacGregor, *loc. cit.*). For work-hardened materials the whole of the curve above the elastic limit is often approximately linear (*cf.* figure 1), and we can therefore write in this case

$$f(\epsilon) = A\epsilon.$$

We then have, for the second term in (10), which represents the contribution to the pressure from the cold work,

$$\frac{4A}{3} \int_a^c \log \left[\frac{r}{\sqrt{(r^2 - a^2)}} \right] \frac{dr}{r}.$$

This reduces to

$$-\frac{A}{3} \left[\sum_1^\infty \frac{1}{n^2 t^{2n}} \right]_1^{c/a}$$

and, if terms in a^2/c^2 are neglected in comparison with unity, to

$$A\pi^2/18.$$

In the general case where the strain-hardening curve is not linear, we shall suppose that above some value ϵ_0 of the strain

$$f(\epsilon) = Y' + A\epsilon; \quad \epsilon > \epsilon_0.$$

A short calculation gives for the second term in (10)

$$\frac{A\pi^2}{18} + \frac{Y'y_0}{\sqrt{3}} \tau \frac{1}{\sqrt{3}} \int_{y_0}^{2 \log(e\epsilon)} \left[f \left\{ \frac{1}{\sqrt{3}} \log \frac{e^y}{e^y - 1} \right\} - \frac{A}{\sqrt{3}} \log \frac{e^y}{e^y - 1} \right] dy$$

where

$$y_0 = \sqrt{3}\epsilon_0 - \log(e^{\sqrt{3}\epsilon_0} - 1).$$

In a similar way we find for the spherical hole the contribution to p_s from the strain hardening is

$$\frac{2A\pi^2}{27} + \frac{2Y'y_0}{3} + \frac{2}{3} \int_{y_0}^{2 \log(e\epsilon)} \left[f \left\{ \frac{2}{3} \log \frac{e^y}{e^y - 1} \right\} - \frac{2A}{3} \log \frac{e^y}{e^y - 1} \right] dy$$

where

$$y_0 = \frac{2}{3}\epsilon_0 - \log(e^{2\epsilon_0/3} - 1).$$

§ 6. NUMERICAL VALUES

We thus see that for a material for which true stress σ and natural strain ϵ are connected by the relation

$$\sigma = Y + A\epsilon,$$

the pressure p_c to enlarge a cylindrical hole is given by

$$p_c = \frac{Y}{\sqrt{3}} \left(1 + 2 \log \frac{c}{a} \right) + \frac{\pi^2}{18} A,$$

with

$$\frac{c}{a} = \left(\frac{\sqrt{3} E}{2(1+\nu)Y} \right)^{\frac{1}{2}}. \quad \dots (13)$$

and for a spherical hole

$$p_s = \frac{2Y}{3} \left(1 + 3 \log \frac{c}{a} \right) + \frac{2\pi^2}{27} A,$$

with

$$\frac{c}{a} = \left(\frac{E}{(1+\nu)Y} \right)^{\frac{1}{3}}.$$

Numerical values are as follows: a typical stress-strain curve for mild steel is shown in figure 7. For this material $(c/a)_{\text{cylinder}} = 22$, $(c/a)_{\text{sphere}} = 8$. In spite of this wide difference, p_c and p_s are nearly equal; in tons/sq. inch,

$$p_c = 100, \quad p_s = 120,$$

the contributions from work-hardening being 29 and 30% in the two cases. Y was taken to be 17 tons/sq. in.

For the cold-worked copper used in the experiments described in sections 2 and 3, similar results are valid, only here the contribution from cold-work is smaller. We took

$$Y = 17.5, \quad A = 6.5, \quad E = 8000,$$

all in tons/sq. inch. ν was not measured, but was assumed to be 0.34. Calculation then gave

$$p_c = 71.3 \quad (3.6 \text{ due to cold work}),$$

$$p_s = 84.5 \quad (4.8 \text{ due to cold work}).$$

It will be noticed again that the values do not differ much. The values of c/a were 17 and 7 respectively.

For the annealed copper shown in figure 2, we found $p_c=43$ and $p_s=52$ tons/sq. in.

§7 COMPARISON WITH STATIC PUNCHING

As already emphasized, we expect the pressure on the head of a punch several diameters deep in the hardened copper to lie between p_c , representing a very sharp punch, and some value below p_s , representing a blunt one (provided friction is eliminated). As explained in section 3, we do not consider the allowance made for friction sufficiently certain to distinguish experimentally between sharp and blunt punches. For a lubricated 20° semi-angle cone our measured value was 85 tons/sq. in., which lies close to our calculated value of P_s , namely 84.5.

For annealed copper, our calculated pressure (between 43 and 52) comes out considerably less than that measured for an unlubricated punch (84 tons/sq. inch). If we ascribe this to a frictional force s per unit area, and take 52 to be the pressure for a lubricated punch, then $52 + s \times 1.73 = 84$, giving $s = 18.5$ tons/sq. inch, which is about the same as we found for hardened copper. Experiments with rotated punches in annealed materials would be of interest. It is perhaps to be expected that the frictional forces on annealed and hardened copper would be about equal to the shear stress of the latter, since the copper near the surface of contact must be highly cold-worked.

§8. SURFACE EFFECTS IN PUNCHING

As figures 3, 4 and 5 show, the load on a punch does not reach its maximum value as soon as the head is embedded in the material, but continues to rise while the penetration increases by several diameters. The same is true of the effective mean pressure on the punch, defined as load divided by the momentary area of the hole on the front surface. The graphs of mean pressure against penetration for the rotated and unrotated 20° S.A. cone are shown in figure 6. On a dimensional argument the pressure will be constant while the conical head is going in. The experimental results do not give such a constant pressure, and the discrepancy is due to an error in the zero measurement of penetration—an error which is proportionately less as the penetration increases. For this reason the straight-line part of the curve is an extrapolation from the results after the head is entirely in the copper.*

It is sometimes stated that the increase in pressure after entry of the head is due only to the strain-hardening of the metal; this is certainly not true, and would take place if the material did not work-harden at all. The region which is plastic round the head of a punch deep in a material must be very extensive, of width between 7 and 17 punch diameters in the case of copper according to the arguments of section 4. Near the surface, however, the constraint on the copper is less, and it can escape by forming a lip round the hole, which is very marked with lubricated punches. Less work per unit volume is therefore done in making the hole. Clearly the small contribution to the maximum pressure by the cold work, calculated in section 6, by no means accounts for the rise in mean pressure from 45 to 85 tons/sq. in. for the rotated 20° punch.

* This constant pressure for the rotated cone is in close agreement with the value obtained by lubrication and repeated loading. See section 9 and figure 8.

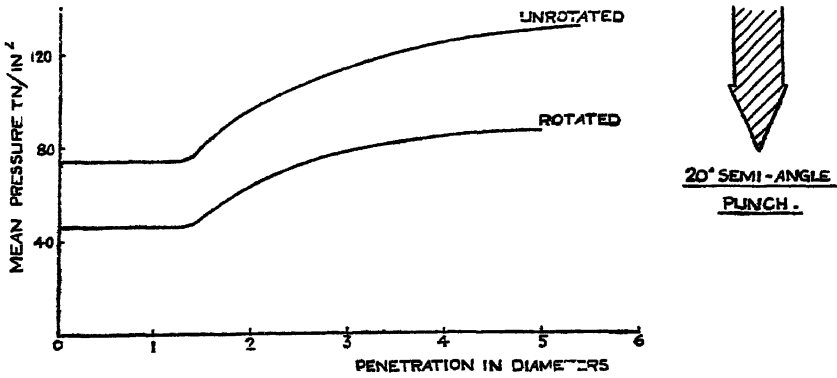


Figure 6 Pressure-penetration curves for 20° semi-angle cone punching into hardened copper.

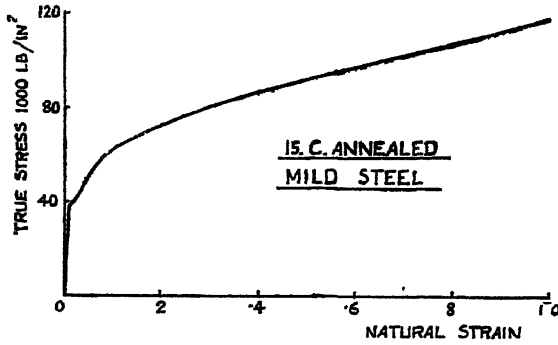


Figure 7.

(Taken from a paper by C W MacGregor in the Timoshenko Anniversary Volume)

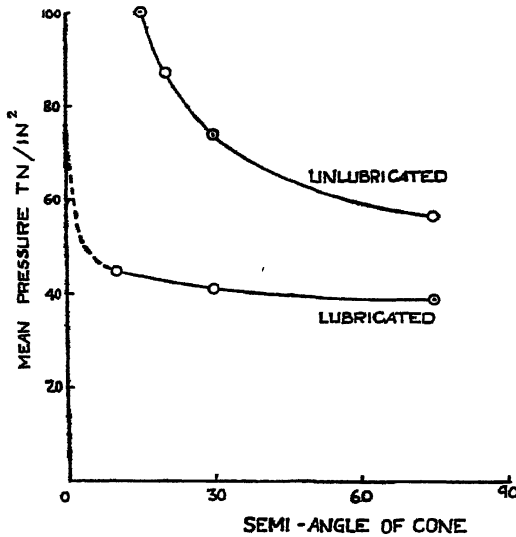


Figure 8. Indentation tests with cones on hardened copper.

Hardness contours taken over a cross-section of a block of annealed copper after punching are shown in figure 9. They imply a smaller diameter of plastic flow near the surface than in the body of the block, as is expected from the

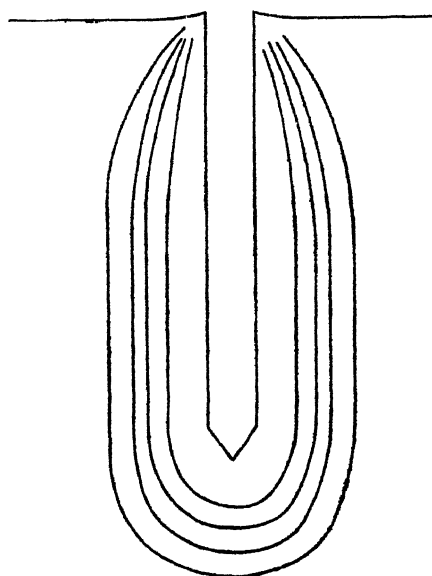


Figure 9 Hardness contours after punching into annealed copper.

discussion given above. Moreover, the actual extent of plastic flow when the punch is beginning to penetrate is still less, since some of the flow near the surface is caused during the later stages of the penetration.

§9 INDENTATION HARDNESS TESTS

Hardness tests with conical indenting tools are particularly suitable for theoretical discussion, because the ratio of load to surface impression area must, on dimensional grounds, be independent of load. Tests with lubricated punches are the simplest to interpret. Such tests have been made by various authors (*cf.* Hankins, 1925 and 1926). In figure 8 we show some results obtained in the Cambridge Engineering Laboratory with lubricated and unlubricated conical punches in strain-hardened copper (with a stress-strain curve as in figure 1). The punches were solid steel cones and the lubricant (graphite grease) was applied repeatedly as the load was increased; at a given load, lubrication was continued until there was no further increase in the diameter of the indentation.

It will be seen that for the lubricated punches, the mean pressure defined by $\text{load}/(\text{surface area of indentation})$ increases slightly with decreasing cone angle, and is about 40 tons/sq. inch. (The opposite is the case when the head of the punch is far from the surface, the mean pressure decreasing with decreasing angle). We conclude that for this material the hardness measured with a lubricated conical punch depends little on the head shape and is about half the experimental ultimate pressure for deep punching (85 tons/sq. in.). Whether this is so for annealed materials will be discussed below.

We consider, on theoretical grounds, that for a cone of sufficiently small angle, the resisting pressure must tend to some value near that for a cylindrical punch or, more exactly, to the theoretical value p_c for the enlargement of a long cylindrical hole. In figure 8, therefore, we have extrapolated the experimental curve for lubricated cones to 71 tons/sq. inch for zero semi-angle (the theoretical value of p_c calculated in section 6). It is to be expected that only for cones of semi-angle less than 10° will these high values be reached. The explanation is to be found in the variation of pressure along the cone, increasing towards the tip, where the confinement is most severe. For example, even a 10° S.A. cone has a penetration of only between 3 and 4 times the impression diameter, whereas we know from deep punching that the maximum pressure is not attained until after 4 diameters penetration. Experiments with sharp cones in a material soft enough to insert them without breaking would be of interest.

The variation in mean pressure with cone angle, small though it is, is an important consideration in the type of experiment where different cones are used to evaluate the coefficient of friction. The calculation based on the experimental values usually involves the subtraction of nearly equal quantities, and so neglect to allow for change of pressure with angle can give completely false results for the friction.

For an indenter with a shape such that the ratio (impression diameter)/(penetration) decreases with depth, the mean pressure steadily increases with penetration for the reasons discussed in section 8. This will be so, for example, in the Brinell test with spheres (where by pressure is meant, as usual, load/(impression area), not load/(spherical area)). This is confirmed experimentally by the empirical result usually referred to as Meyer's law. This states that the load L on a ball is proportional to d^n , where d is the impression diameter and n is greater than 2. According to the ideas of this paper, n should be slightly greater than 2 even for a material which does not show work-hardening.

Empirically n is largest for annealed materials and least for severely strain-hardened materials. It is clear from general considerations that this should be so, as pointed out in various papers at a discussion held by the Institution of Mechanical Engineers. For a shallow indentation by a sphere, or an indentation by a very blunt cone, no part of the material will be highly strained. For a deep penetration by a sphere, or penetration by a sharp cone, large strains will be involved; work-hardening will thus take place.

We therefore expect that for an annealed material the pressure on a lubricated conical punch would rise more rapidly with decreasing cone angle than shown in figure 8. While (in terms of the yield stress) the value for the 90° cone would be unaffected, that for the 0° cone would be increased by work-hardening as calculated in formulae (12). Experiments with an annealed material would be of interest.

As regards the limiting pressure for blunt cones ($\theta = 90^\circ$) shown in figure 8 it is of interest to compare this with the theoretical value obtained by Hencky (1923) by a very approximate method, according to which yield with a cylindrical flat-headed punch should begin when $p = 2.8Y$. Since Y was 17.5 tons/sq. inch, Hencky's formula gives 49 against the observed value of 39 tons/sq. inch.

Various empirical rules have been given from time to time in order to estimate the tensile yield stress from a hardness test. It is clear that no general conversion factor can be laid down, even for the same material, since the factor will depend largely on the amount of pre-working. But for severely cold-worked metals there seem to be good reasons for supposing a constant conversion factor to exist independent of the material tested, with the proviso that the impression diameter must be the same in every case. In this way it is possible to eliminate the effect of variation of mean pressure with penetration, since the variation will then be the same for all the materials tested. A second precaution is the elimination of friction. The only variables on which the conversion factor can now depend are the yield stress and elastic constants (for the same indenter). The elastic properties are probably of not more than minor importance since we know that in a cold-worked metal the material escapes from the indenter into the lip.

In these circumstances the mean pressure will probably be between two and three times the yield stress of the worked metal, depending on the geometry of the indentation. For the cones indenting the hardened copper of yield strength 17.5 tons/sq. inch the pressure was about 40 tons/sq. inch, giving a ratio of 2.3.

ACKNOWLEDGMENTS

The theoretical work of this paper was carried out in the Branch for Theoretical Research of the Armament Research Dept., and the experimental work in the Engineering Laboratory, Cambridge University. The authors are indebted to the Director-General of Scientific Research and Development, Ministry of Supply, for permission to publish this paper.

REFERENCES

- HANKINS, G. A., 1925. *Proc. Inst. Mech. Engrs.* 611, 1926. *Ibid.* 823.
HENCKY, H., 1923. *Z. angew. Math. Mech.* 3, 250.
INST. MECH. ENGRS., 1944. Discussion of the significance of the tensile and other mechanical test properties of metals. *Proc. Inst. Mech. Engrs.* 151, 105 *et seq.*
MACGREGOR, C. W., 1938. Timoshenko Anniversary Volume. Article on "Differential Area Relations in the Plastic State".
NADAI, A., 1937. *J. Appl. Phys.* 8, 205.

AN EXPERIMENTAL INVESTIGATION OF EXTRAPOLATION METHODS IN THE DERIVATION OF ACCURATE UNIT-CELL DIMENSIONS OF CRYSTALS

BY J. B. NELSON AND D. P. RILEY,
Cavendish Laboratory, Cambridge

MS. received 6 December 1944

ABSTRACT Measurements on x-ray photographs of cylindrical specimens of different absorption and thickness taken in a camera without eccentricity show that the absorption error in the apparent unit-cell dimension a is proportional to

$$\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta}.$$

The plot of a against $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ is linear down to $\theta = 30^\circ$ for all four specimens used. The extrapolated values for a are in good agreement, and this extrapolation function is accordingly recommended in the case of data from well constructed cameras. Other extrapolation functions are also considered, and the effect of various sources of error discussed. A table of $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ is given.

§ 1 INTRODUCTION

IN the determination of accurate unit-cell dimensions from powder photographs, three sources of systematic error are usually considered, due to

- (a) absorption of the x-ray beam in the specimen,
- (b) the displacement of the rotation axis of the specimen relative to the geometric centre of the cylindrical film, usually called the *eccentricity error*,
- (c) inaccurate determination of the camera constants.

It may be shown that errors due to these three causes vanish at $\theta = 90^\circ$, which corresponds to extreme back-reflection along the path of the incident beam. The procedure usually adopted, therefore, is to derive apparent cell dimensions a from a number of lines on the photograph, to plot these values against some function of the Bragg angle θ , and to extrapolate to a value corresponding to $\theta = 90^\circ$.

Some uncertainty exists as to the best function of θ to use in order to obtain a linear extrapolation. Current practice is to plot a against $\cos^2 \theta$ for high-angle lines only (i.e. those with clearly resolved α -doublets) and to extrapolate linearly to $\cos^2 \theta = 0$. It appears that this function has been chosen because the eccentricity error $\left(\frac{\delta a}{a} \right)_{\text{ecc}}$ is easily shown to be proportional to $\cos^2 \theta$. The absorption

error $\left(\frac{\delta a}{a}\right)_{\text{abs}}$, however, is not proportional to $\cos^2 \theta$. Various workers, on theoretical grounds, have derived different functions relating $\left(\frac{\delta a}{a}\right)_{\text{abs}}$ and θ . Thus

1. $\left(\frac{\delta a}{a}\right)_{\text{abs}} \propto \frac{\cos^2 \theta}{\theta}$ Bradley and Jay (1932).
2. $\left(\frac{\delta a}{a}\right)_{\text{abs}} \propto \frac{\cos^2 \theta}{\sin \theta}$ Jay (1944).
3. $\left(\frac{\delta a}{a}\right)_{\text{abs}} \propto \cot \theta \cos^2 \theta$ Buerger (1942).
4. $\left(\frac{\delta a}{a}\right)_{\text{abs}} \propto \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta}\right)$ Taylor and Sinclair (1944).*

The use of a $\cos^2 \theta$ plot assumes that, whatever the nature of the absorption error function, it can be approximately represented by $\cos^2 \theta$ at high angles. It is not always possible, however, to obtain a sufficient number of lines at angles high enough for this to be true, and we have carried out the work described here in order to obtain experimental evidence concerning the best extrapolation function for general use.

§ 2. EXPERIMENTAL

The camera used was the 19-cm. diameter high-temperature camera designed and described by Wilson (1941), the film being in two halves. The design minimizes the possibility of eccentricity, and after consultation with Mr. C. E. Chapman, who made the instrument, we satisfied ourselves that effectively no eccentricity could be present. The camera angle, θ_k , has been very carefully measured by Wilson and Lipson (1941), who discussed possible errors in this connection. In effect, then, we have used an excellently constructed and calibrated camera to investigate the *absorption error*, eccentricity and calibration errors being absent. The method of film measurement described by Lipson and Wilson (1941) avoids errors due to uniform film shrinkage, but requires an accurate value for the camera angle. We therefore thought it instructive to do some calculations using camera angles in error by a known amount.

For simplicity, the substance photographed should be cubic and give a large number of reasonably strong and sharp lines over a wide range of θ values. We chose the γ -structure Cu_2Al_3 , which, with Cu radiation, gives a well resolved and intense α -doublet ($h^2 + k^2 + l^2 = 126$) at $\theta_{\alpha} = 83^\circ.4 : \theta_{\alpha} = 84^\circ.8$, together with a large number of lines going down to the lowest order line measurable ($h^2 + k^2 + l^2 = 6$) with $\theta = 12^\circ.7$. Table 1 gives a list of the lines measured, which will be seen to be distributed fairly evenly over the photograph. The Cu radiation was filtered to avoid the confusion inevitable with β -lines in a pattern of this complexity. The specimens were accurately centred and rotated during the exposure. The temperature for each exposure was kept constant by circulating water through the cooling jacket of the camera. Care was taken to ensure that the electron beam in the Metropolitan-Vickers x-ray tube was homogeneously

* Subsequently to our work.

Table 1
X-ray reflections from Cu_3Al_4 ; $\text{CuK}\alpha$ radiation
(Approximate θ values from thick diluted specimen)

$h^2 + k^2 + l^2$	θ	$h^2 + k^2 + l^2$	θ	$h^2 + k^2 + l^2$	θ
6	12.8	66 a_1	46.2	102 a_2	63.8
9	15.7	66 a_2	46.3	108 a_1	67.0
12	18.2	72 a_1	48.9	108 a_2	67.3
14	19.7	72 a_2	49.0	114 a_1	71.0
18	22.4	76 a_1	50.7	114 a_2	71.4
22	24.8	78 a_1	51.6	118 a_1	74.1
24	26.0	90 a_1	57.3	120 a_1	75.9
36 a_1	32.3	90 a_2	57.5	120 a_2	76.4
48 a_1	38.1	98 a_1	61.3	122 a_1	77.9
54 a_1	40.8	98 a_2	61.6	126 a_1	83.4
54 a_2	40.9	102 a_1	63.5	126 a_2	84.8

distributed over the target, i.e. no "hot-spots" were present. The camera was adjusted so that its slit-system made a small angle with the face of the target, thus securing a "foreshortened" circular focal spot, and this arrangement was accurately reproduced for each exposure.

§ 3. SPECIMENS USED

Filings of Cu_3Al_4 were annealed at 650° for $\frac{1}{2}$ hour and sieved through a 350 B.S.S. mesh. This sample was used to prepare specimens (a) and (b) described below. Specimens (c) and (d) were prepared from the same filings annealed at a slightly higher temperature, 700° , for $\frac{1}{2}$ hour.

Four specimens were prepared :

- (a) A rod 0.59 mm. diameter made by rolling with gum tragacanth.
- (b) As above, 1.46 mm. diameter.
- (c) A silica tube, bore 0.45 mm., uniform wall thickness 0.025 mm., filled with Cu_3Al_4 powder packed as densely as possible.
- (d) A Lindemann-glass tube, bore 1.35 mm., uniform wall thickness 0.016 mm., filled as above.

The use of gum tragacanth may be unfamiliar and will be described. The specimens are made by mixing the specimen powder with gum tragacanth powder, moistening, and rolling the resulting dough into a uniform rod. When dry, this rod is quite rigid. Specimen (a) was the thinnest one it was convenient to make with this technique, and was prepared with enough gum tragacanth roughly to satisfy the generally accepted condition for ensuring good photographs, namely, that $\mu r = 1$, where μ = linear absorption coefficient of the specimen mixture, r = radius of specimen, the gum tragacanth acting as a diluent. Specimen (b) was prepared from the same mixture but was deliberately made very thick; the condition $\mu r = 1$, of course, no longer applied. Specimens (c) and (d) approximated in density and absorption to a wire specimen. Here again one specimen was made as thin as possible and the other very thick. It should be pointed out

that each of our specimens was a homogeneous solid cylinder. The method of preparing specimens by causing the powdered sample to adhere to the surface of a hair or a thin glass fibre gives a hollow cylinder.

§ 4. EXTRAPOLATIONS

The results for specimen (b) (the thick gum tragacanth specimen) were considered in some detail. The apparent a values were plotted against the following functions of θ :

$$\theta^\circ, \cot \theta, \cos^2 \theta, \cot \theta \cdot \cos^2 \theta, \cot \theta \left(\frac{1 + \cos^2 \theta}{2} \right),$$

$$\frac{\cos^2 \theta}{\sin \theta}, \quad \frac{\cos^2 \theta}{\theta}, \quad \frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right).$$

These plots are shown in figure 1, and display several interesting features.

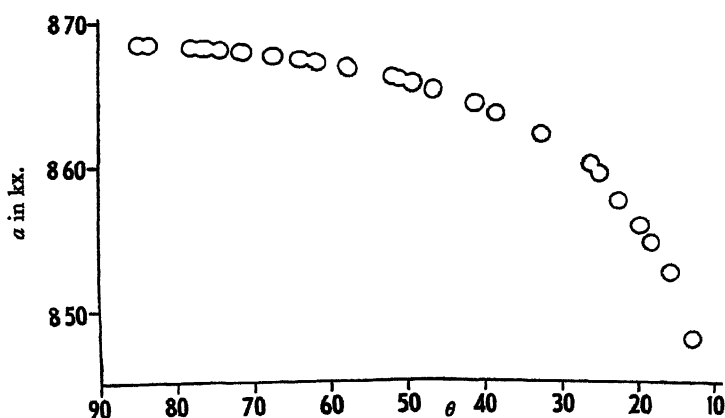


Figure 1 a.

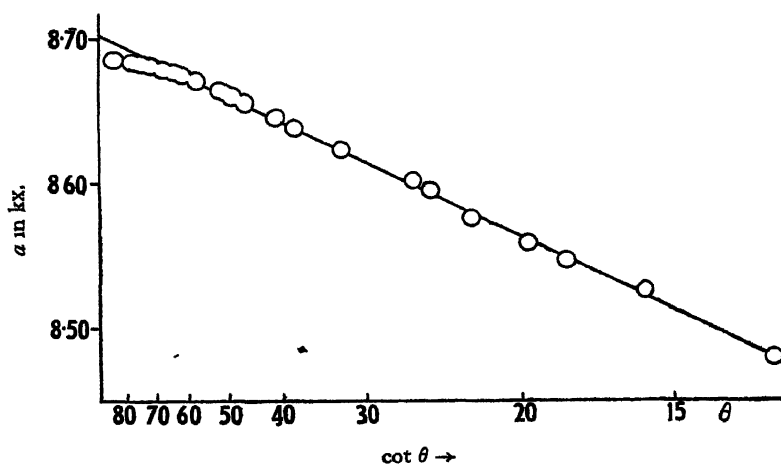


Figure 1 b.

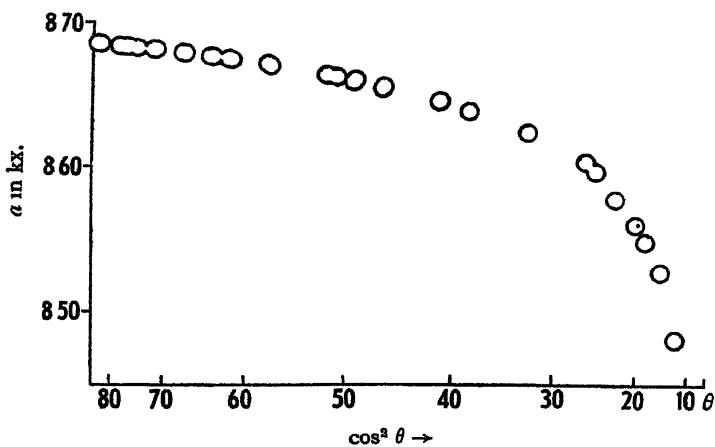


Figure 1 c

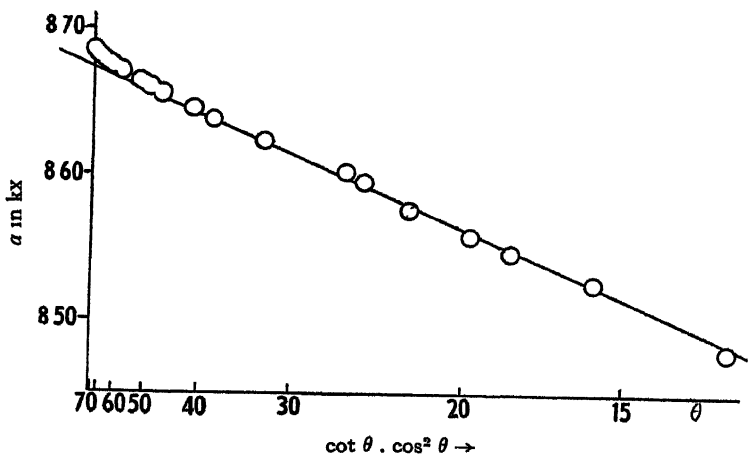
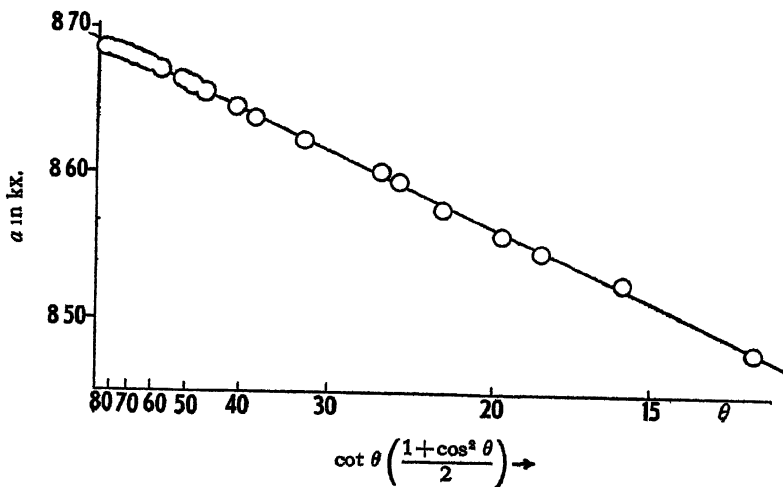


Figure 1 d.



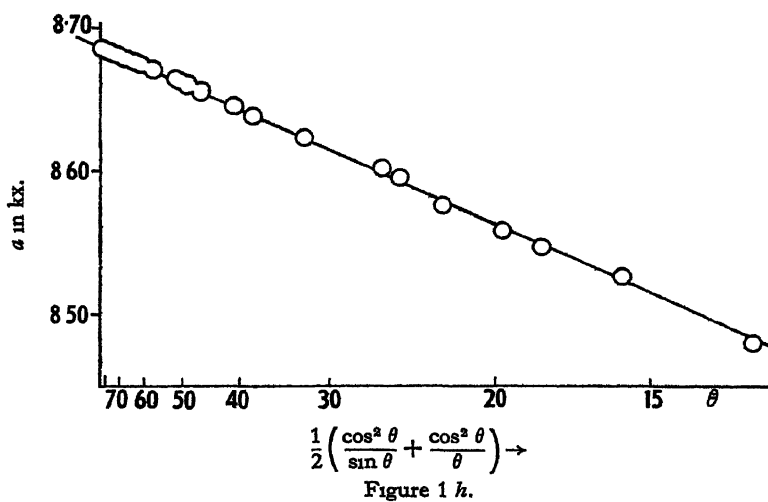
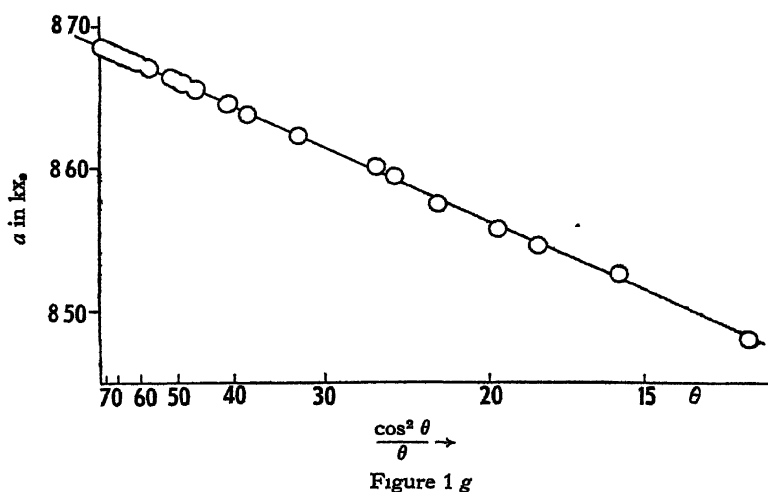
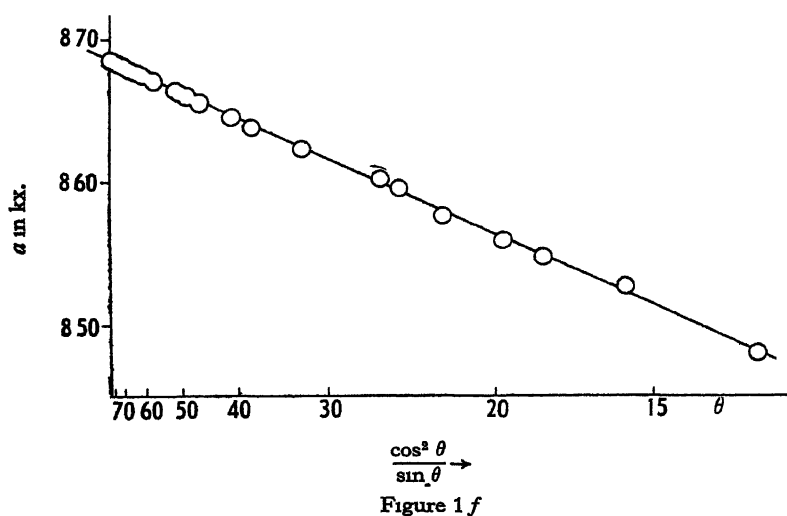


Figure 1 Thick diluted specimen.
Plots of a against various functions of θ .

Firstly, the $\cos^2 \theta$ plot is nearly linear only over a very small range, and its main advantage over the straightforward plot against θ is the way in which it compresses the high angle values towards $\cos^2 \theta = 0$. Secondly, it is clear that the plot must be against $\cot \theta$, or against a very similar function, if linearity over the whole range is to be achieved. The plot against $\cot \theta$ itself shows a falling off from linearity at high angles, while the use of $\cot \theta \cos^2 \theta$ (Buerger's absorption-error plot) gives the opposite deviation. A plot against the arithmetic mean of these two functions, $\cot \theta \frac{1 + \cos^2 \theta}{2}$, shows a marked improvement but still falls off at high angles. The plot against $\frac{\cos^2 \theta}{\sin \theta}$ (or $\cot \theta \cos \theta$, the geometric mean) is apparently linear, as is the plot against the very similar function $\frac{\cos^2 \theta}{\theta}$. A close examination of the two extrapolations, however, shows that in the case of the $\frac{\cos^2 \theta}{\sin \theta}$

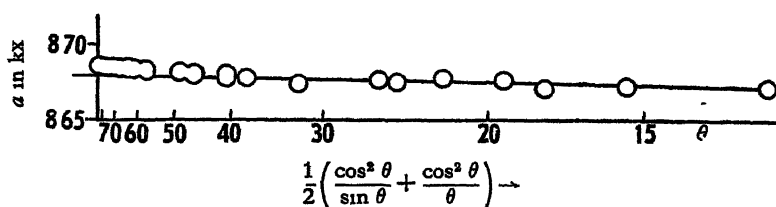


Figure 2 a.

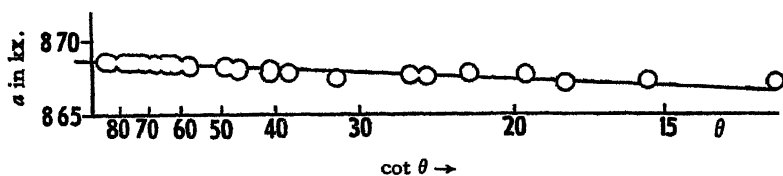


Figure 2 b.

Figure 2. Medium thickness diluted specimen.

plot the best straight line drawn through the high-angle points only has a slightly smaller slope than the best straight line drawn through all the points, whereas the reverse is true for the $\frac{\cos^2 \theta}{\theta}$ plot. In conformity with these observations, a remarkably good linear plot is obtained against the arithmetic mean of these functions, $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$. In this case the best overall extrapolation coincides exactly with the best high-angle extrapolation, as is shown more clearly in the larger scale drawing in figure 5 a (lower plot).

The results for specimen (a), the gum tragacanth specimen of medium thickness, are not in agreement with these findings. The plot against $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ is not linear over the whole range (figure 2 a). The high-angle points fall closely on a straight line, but the slope of this line is markedly greater than that of the line drawn through the low-angle points. The latter extrapolation,

it is interesting to note, is almost horizontal. The plot against $\cot \theta$, on the other hand, is linear over the whole range (figure 2*b*).

The thick undiluted specimen (*d*) gives similar results to those obtained with specimen (*b*). Plotting against $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ gives a remarkably good straight line over the whole range (figure 3).

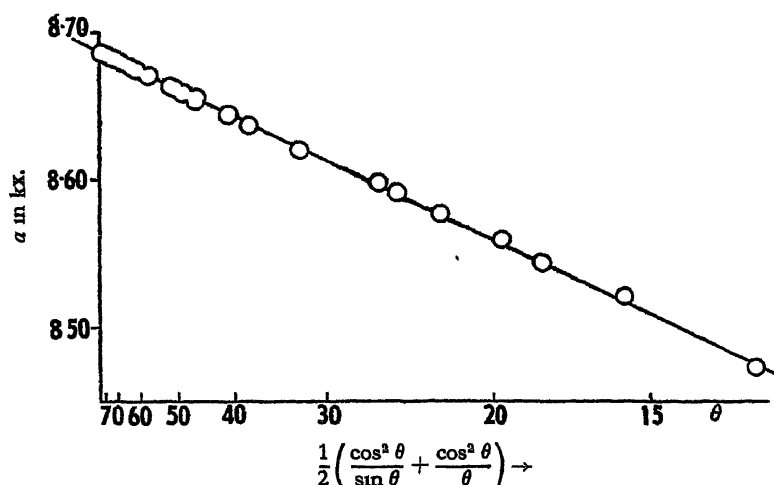


Figure 3 Thick undiluted specimen.

The undiluted specimen of medium thickness, specimen (*c*), behaves similarly to specimen (*a*). The plot against $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ is not linear over the whole range, although a good straight-line extrapolation can be obtained with the high-angle points (figure 4*a*). The linearity of the plot against $\cot \theta$ (figure 4*b*) is not as good as with specimen (*a*), as there is a slight downwards bend at high angles. A plot against the arithmetic mean of the two functions is nearly linear over the whole range, but is unsuitable for extrapolation.

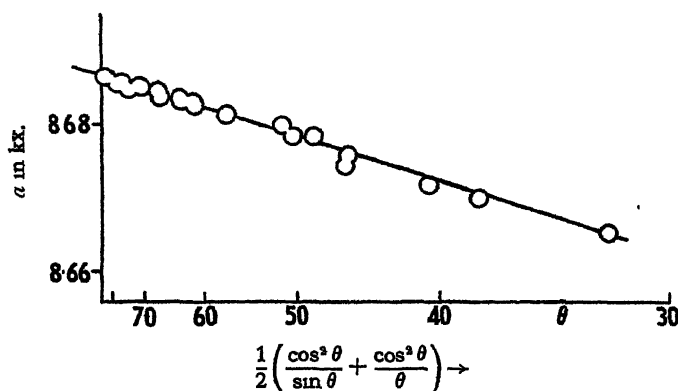


Figure 4 *a*

Large scale plot of a against $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ for lines with $\theta > 30^\circ$.

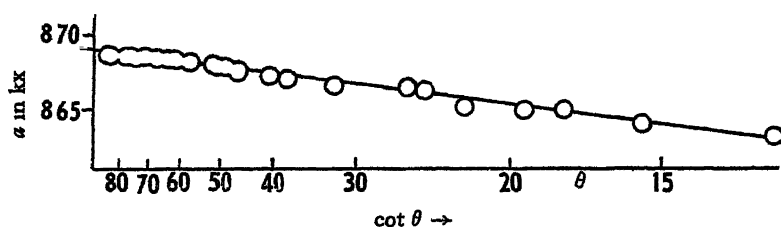


Figure 4 b

Figure 4. Medium thickness undiluted specimen.

These results are summarized in the following table:

Specimen	Best overall plot for linearity	Best high-angle plot for accuracy	Low-angle limit of linearity for high-angle plot (θ)
(a) Medium thickness diluted	$\cot \theta$	$\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$	30°
(b) Very thick diluted	$\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$	"	$(13^\circ)^*$
(c) Medium thickness undiluted	$\frac{1}{2} \left[\cot \theta + \frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right) \right]$	"	30°
(d) Very thick undiluted	$\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$	"	$(13^\circ)^*$

* For line 6, the lowest line measured. The true limit may be lower than this.

We interpret these results in the following way. Two factors are contributing to the errors in the apparent a values, (a) the absorption, and (b) systematic observational errors.

(a) The absorption error

$$\left(\frac{\delta a}{a} \right)_{\text{abs}} \propto \frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta}.$$

This is in agreement with the theoretical finding of Taylor and Sinclair for the case of an "exponential" focus, although we had established the relation experimentally before they had reached this conclusion.

(b) Errors in a due to systematic observational errors in measuring θ are proportional to $\cot \theta$. This is easily confirmed by differentiating the Bragg equation:

$$a = \frac{\lambda}{2} \sqrt{h^2 + k^2 + l^2} \operatorname{cosec} \theta \quad (\text{for a cubic crystal});$$

therefore
$$\delta a = -\frac{\lambda}{2} \sqrt{h^2 + k^2 + l^2} \operatorname{cosec} \theta \cot \theta \delta \theta$$

and
$$\frac{\delta a}{a} = -\cot \theta \delta \theta.$$

Systematic observational errors may arise in measuring the positions of lines on a film using the usual type of measuring instrument. An observer may systematically place the cross-hairs slightly to one side of the position of maximum line density, particularly if the background blackening has a marked slope. We have noticed that there is often a slight systematic discrepancy between readings made independently by the two of us.

In the thick specimens, absorption is the over-riding source of error, while in the specimen of "optimum" absorption and medium thickness, the effect of absorption is very slight. Hence the thick specimens give a linear plot against the absorption error function. When, however, the effect of absorption is negligible, the plot is no longer linear against this function; the only remaining source of error is observational and the plot becomes instead linear against $\cot \theta$. An intermediate case exists for the undiluted specimen of medium thickness. Here the plot is not strictly linear against either the absorption error function or against $\cot \theta$. The effect of observational errors being comparable with the absorption error, the plot is nearly linear against a mean of $\cot \theta$ and $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$.

Having considered the question of the overall linearity of the plots, it is necessary to investigate the consistency with which any given type of plot will lead to a single value for a . Large-scale plots of apparent a vs. $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ for the high-angle lines ($\theta > 30^\circ$) are shown in figures 5 *a* and 5 *b*.

In both cases, the extrapolations for the medium and thick specimens converge very nearly to a point in spite of the considerable difference in slope. The following table gives a list of the extrapolated a values obtained, the maximum extrapolation deviation in each case being ± 0.0001 kx.

Specimen	a from $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ extrapolation			Temp.
(a) Medium diluted	8.6863 kx	} difference = 0.0002 kx.	"	15°·8
(b) Thick diluted	8.6861 kx			15°·4
(c) Medium undiluted	8.6866 kx.			16°·2
(d) Thick undiluted	8.6864 kx			16°·4

Mean value for $a = 8.6864 \pm 0.0003$ kx.

It will be noticed that the medium specimens in each case give a slightly larger value for a than the thick, and that the undiluted specimens give slightly higher values than the diluted. The slightly higher temperatures at which the undiluted specimens were photographed would tend to give slightly higher values for a . It is unlikely that the small difference of annealing temperature would have any measurable effect. Even so, the maximum error in the mean value of a from the four results is only about one part in 30,000. It is of interest that the value of a derived from the medium diluted specimen of "optimum" absorption is very close to the mean value.

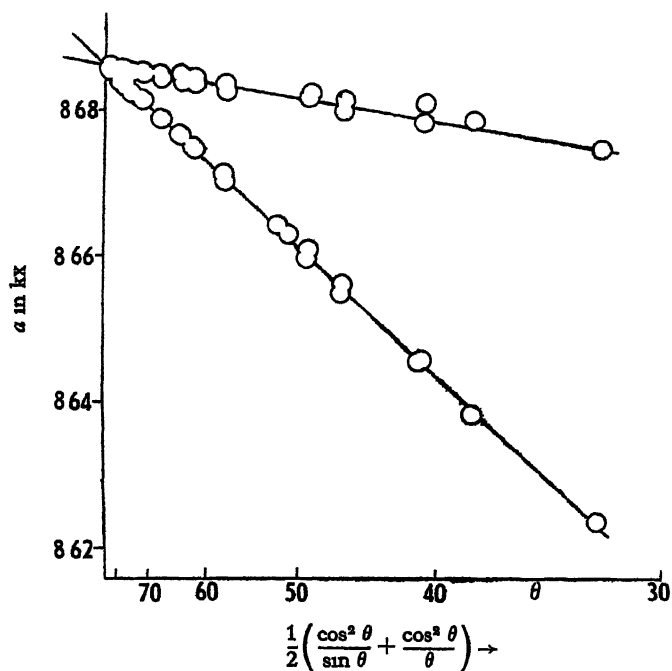


Figure 5 a Medium thickness diluted specimen (upper plot).
Thick diluted specimen (lower plot).

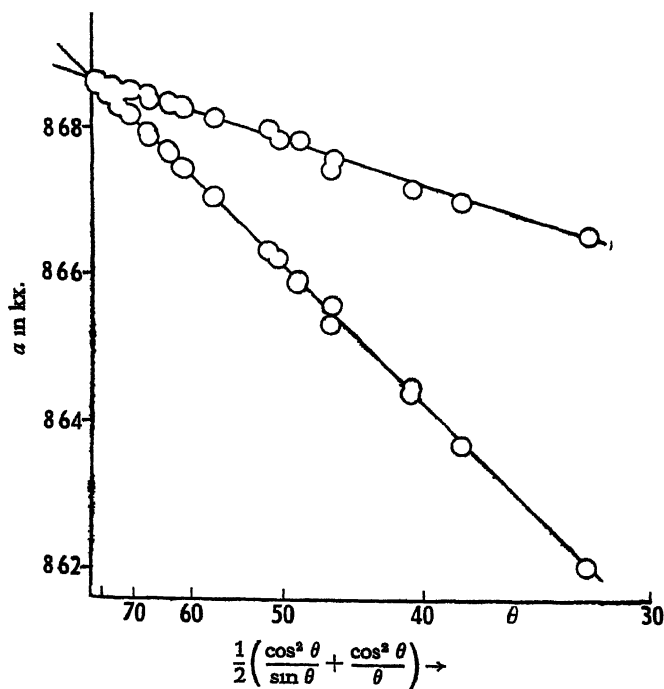


Figure 5 b. Medium thickness undiluted specimen (upper plot).
Thick undiluted specimen (lower plot).

Figure 5. Large-scale plots of a against $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ for lines with $\theta > 30^\circ$.

A similar large-scale plot was made for specimen (a) (diluted, medium thickness) against the function which gave the best overall linearity, namely, $\cot \theta$. This is shown in figure 6.

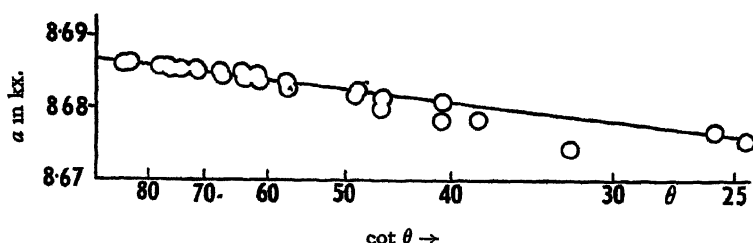


Figure 6 Medium thickness diluted specimen
Large-scale plot of a against $\cot \theta$ for lines with $\theta > 25^\circ$.

The extrapolated value of a is 8.6867 kx., which is higher than that obtained above. It was noticed that the $\cot \theta$ plot tended to emphasize the effect of random errors in measurement, giving a slightly more "scattered" plot than the one against $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$.

The enclosure of specimens in thin-walled silica or Lindemann-glass tubes does not appear to cause any new errors.

§5 EFFECT OF WRONG CAMERA ANGLE

The effect of an error in θ_k , the camera angle, is to cause a flexure in the plot of a vs. $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ at high angles. Figures 7a and 7b show this effect on data obtained from specimen (b). The a values used were calculated using θ_k in error by 0.1% in a positive and negative sense. A positive error in θ_k causes an upward flexure, and *vice versa*. The symmetry of the two effects confirms the accuracy of determination of the camera angle ($\theta = 86^\circ 6' 93''$) by Wilson and Lipson (1941).

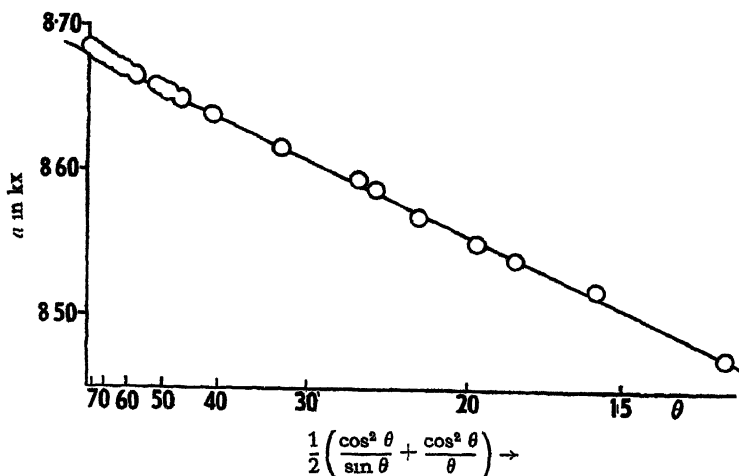
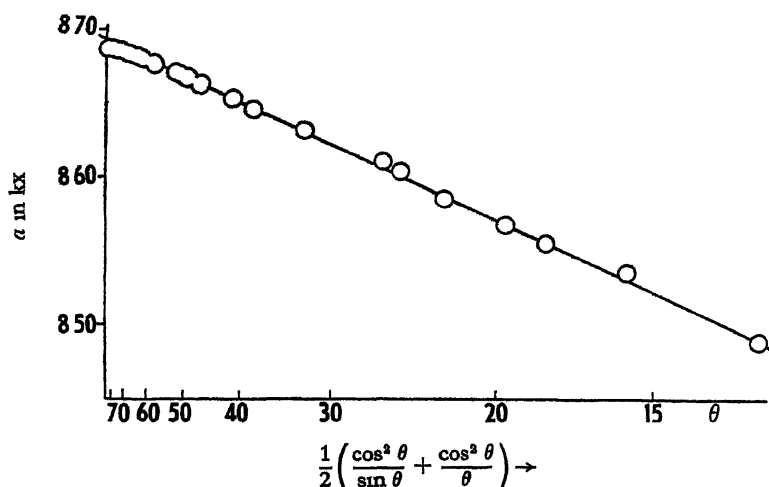


Figure 7a. θ_k in error by +0.1%

Figure 7b. θ_k in error by -0.1% Figure 7. Effect of wrong camera angle, θ_k .

Plots of a against $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ for thick diluted specimen.

These authors also discussed the source of error and showed that errors in a due to this cause are proportional to $\theta \cot \theta$. It is advisable to use a camera which can record lines with θ approaching 85° , otherwise the flexure in the plot may be unobservable and an unsuspected source of error be present.

§ 6. METHOD OF CALIBRATING CAMERA

The recommendation of Wilson and Lipson (1941) is to calibrate by direct measurement. Some cameras are so constructed, however, that this is impossible, and θ_k must then be derived from the high-angle lines on a quartz photograph. The basis of the method as described by Bradley and Jay (1933) is to derive values of θ_k from each high-order reflection and then to use an extrapolation technique. These authors showed that,

$$\text{if } \frac{\delta a}{a} \propto \cos^2 \theta,$$

$$\text{then } \frac{\delta \theta_k}{\theta_k} \propto \frac{\sin 2\theta}{2\theta}.$$

Hence the true value of θ_k is found by plotting the apparent values against $\frac{\sin 2\theta}{2\theta}$ and extrapolating linearly to zero (corresponding to $\theta = 90^\circ$).

We have shown that with well-constructed cameras absorption is the main source of error, and that

$$\delta a \propto \frac{\cos^2 \theta}{\sin \theta} \cdot \frac{\cos^2 \theta}{\theta}$$

In this case it is easily proved that

$$\frac{\delta \theta_k}{\theta_k} \propto \frac{\cos \theta}{\theta} \left(1 + \frac{\sin \theta}{\theta} \right).$$

Hence a plot of θ_k against $\frac{\cos \theta}{\theta} \left(1 + \frac{\sin \theta}{\theta}\right)$ should give a straight line, and extrapolation to zero value of this function (corresponding to $\theta = 90^\circ$) should give the correct value of θ_k . Table 2 gives the values of $\frac{\cos \theta}{\theta} \left(1 + \frac{\sin \theta}{\theta}\right)$ for the angles of reflection for quartz recommended by Wilson and Lipson (1941).

We have tried out this method of extrapolation on the results of Wilson and Lipson for films 873, 876, 877, which were all taken with the camera we used. The values derived for θ_k were $86^\circ 687$, $86^\circ 694$, $86^\circ 704$; mean $\theta_k = 86^\circ 695$,

Table 2

Values of $\frac{\cos \theta}{\theta} \left(1 + \frac{\sin \theta}{\theta}\right)$ for quartz. Cu radiation, 18° c.

Line	ν ()	$\frac{\cos \theta}{\theta} \left(1 + \frac{\sin \theta}{\theta}\right)$	Line	ν	$\frac{\cos \theta}{\theta} \left(1 + \frac{\sin \theta}{\theta}\right)$
412 a_1	61.304	0.817	502 a_1	71.632	0.443
412 a_2	61.567	0.806	225 a_1		
305 a_1	63.631	0.723	502 a_2	72.065	0.430
305 a_2	63.920	0.711	225 a_2		
			331 a_1		
116 a_1	65.613	0.647			
116 a_2	65.931	0.636	331 a_2	72.509	0.417
501 a_1	very faint		420 a_1	73.319	0.392
501 a_2			420 a_2	73.803*	0.378
404 a_1			315 a_1	75.092*	0.341
404 a_2					
206 a_1	68.217	0.555	315 a_2	75.606	0.327
206 a_2	68.578	0.542	421 a_1		
			421 a_2	76.138	0.312
413 a_1	68.951	0.530	234 a_1	76.772	0.295
413 a_2	69.326	0.517	234 a_2	77.394	0.278
330 a_1	70.162	0.490	216 a_1	78.552	0.248
330 a_2	70.562	0.477	216 a_2	79.280	0.230

* These angles are given erroneously in table 4 of Wilson and Lipson's paper (1941), owing to a proof error.

mean deviation = $0^\circ 006$. The above authors, using a $\frac{\sin 2\theta}{2\theta}$ plot, had obtained $86^\circ 688$, $86^\circ 697$, $86^\circ 709$; mean $\theta_k = 86^\circ 698$, mean deviation = $0^\circ 007$. The camera angle derived by direct measurement is $86^\circ 693$, which is taken as the correct value. It will therefore be seen that the $\frac{\cos \theta}{\theta} \left(1 + \frac{\sin \theta}{\theta}\right)$ plot gives a more accurate result for θ_k in this case.

We should like to emphasize the recommendations of Wilson and Lipson that cameras should be calibrated by direct measurement wherever possible. When quartz has to be used, we would advise that at least three photographs be taken.

§ 7. RECOMMENDATIONS ON PROCEDURE

It is obvious that if a number of factors are contributing to the error, each of the factors having a different dependence on θ , no satisfactory linear plot of a vs. $f(\theta)$ can in general be obtained. The eccentricity error can, however, be almost entirely avoided by using a well-made camera, which should be a *sine qua non* in an investigation claiming the highest accuracy. The principal remaining source of error is then due to absorption, and linear extrapolation over a wide range becomes possible. It may not always be easy to assess the importance of eccentricity in the camera being used. In such cases it is advisable to calibrate the camera for extrapolation, following a procedure similar to that described in this paper. That is to say, the extrapolation function best suited to the particular camera should be derived empirically.

The first general recommendation, then, is to minimize the possible sources of error by using a well-designed and accurately constructed camera. Secondly, whether the Bradley-Jay or van Arkel arrangement is used, the method of film measurement using a camera angle described by Bradley and Jay (1932) and by Lipson and Wilson (1941) should be employed as it avoids uniform film-shrinkage errors. It is necessary, however, to know the camera angle accurately. We have shown in this paper that the effect of 0.1% error in the camera angle is easily noticed in an extrapolation, and Wilson and Lipson observed it with an error of only 0.06%. Hence the linearity of the plot at very high angles should provide another check on the accuracy of calibration of the camera. Thirdly, care should be taken to avoid "hot spots" on the target and to use an effectively circular focal spot.

Four conditions need to be considered in deciding on an extrapolation procedure :

- (a) It should give a consistent value for a whatever the nature of the specimen.
- (b) The low-angle limit of linearity of the plot should be as low as possible.
- (c) The compression of the high-angle points towards the extrapolation limit should be considerable.
- (d) The slope of the extrapolation should be as small as possible.

It is clear from the above results that a thick absorbing specimen gives an accurate linear plot of a vs. $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ down to the lowest recorded values of θ . This advantage is somewhat outweighed by the greatly increased slope of the extrapolation. A thinner specimen of "optimum" absorption will, on the other hand, give rise to a very small extrapolation slope. The plot in general will not be linear over the whole range, but a plot of a vs. $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ will be accurately linear down to an angle of about 30° . This is a very much lower limit than is possible with a plot against $\cos^2 \theta$. Although a plot against $\cot \theta$ for such a specimen may have a greater range of linearity, it has the disadvantage that the high-angle points are not brought very close to the extrapolation limit. It also tends to accentuate the effect of random errors. Further, it has been

demonstrated for such a specimen that the plot against $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ gives a value for a very close to the mean obtained from a number of specimens.

In order to obtain highest accuracy it is therefore recommended.

(a) To make a thin specimen of "optimum" absorption.

(b) To choose x radiation of the right wave-length to give high-angle lines; in particular, to ensure that at least one line has $\theta > 80^\circ$. The closer this final line is to $\theta = 90^\circ$, the greater the accuracy, no matter what extrapolation method is used.

(c) To plot apparent a against $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ and to extrapolate linearly to zero value of this function. Lines with θ values as low as 30° can be included in the plot. This means that a far greater number of lines than usually considered can justifiably be used, giving the extrapolation greater certainty.

In critical cases it may be advisable to make a number of specimens of different thickness and absorption and to take the mean value of the results.

It must be emphasized that these recommendations have effect only when the eccentricity error is negligible, and we have urged that cameras be so designed and constructed that this is so. If, however, the effect of eccentricity is very much greater than that of absorption, a plot against $\cos^2 \theta$ would be more correct. It is our experience that this is rarely the case, i.e. that absorption is usually the more important source of error.

If the problem is such that a reasonably accurate value of the cell dimension has to be derived from a few low-angle lines only, it would be preferable to make a thick absorbing specimen and again to use the $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ plot. The reason for this is that the extrapolation will have greater certainty than would be the case for the low-angle lines of a thin specimen. At the same time, it should be borne in mind that a thick specimen causes a considerable diminution in intensity of low-angle lines.

Our recommendations may be criticized on the grounds that they are based on the examination of one substance only, Cu_3Al_4 . There is, however, no reason for supposing any uniqueness of behaviour for this substance in the derivation of accurate unit-cell dimensions. Also, the different types of specimens prepared were chosen so as to give a considerable variation in properties.

A table of $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$ calculated for every tenth of a degree is appended (table 3). The calculations in this paper were carried out on a Marchant electric calculating machine with automatic division.

ACKNOWLEDGMENTS

We are indebted to Dr. A. J. Bradley, F.R.S., for the sample of Cu_3Al_4 used, and to Drs. H. Lipson and A. J. C. Wilson for several helpful discussions. The work was done in connection with an investigation on graphite carried out for the British Coal Utilisation Research Association.

Table 3
Table of $\frac{1}{2} \left(\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta} \right)$

θ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10	5.572	5.513	5.456	5.400	5.345	5.291	5.237	5.185	5.134	5.084
11	5.034	4.986	4.939	4.892	4.846	4.800	4.756	4.712	4.669	4.627
12	4.585	4.544	4.504	4.464	4.425	4.386	4.348	4.311	4.274	4.238
13	4.202	4.167	4.133	4.098	4.065	4.032	3.999	3.967	3.935	3.903
14	3.872	3.842	3.812	3.782	3.753	3.724	3.695	3.667	3.639	3.612
15	3.584	3.558	3.531	3.505	3.479	3.454	3.429	3.404	3.379	3.355
16	3.331	3.307	3.284	3.260	3.237	3.215	3.192	3.170	3.148	3.127
17	3.105	3.084	3.063	3.042	3.022	3.001	2.981	2.962	2.942	2.922
18	2.903	2.884	2.865	2.847	2.828	2.810	2.792	2.774	2.756	2.738
19	2.921	2.704	2.687	2.670	2.653	2.636	2.620	2.604	2.588	2.572
20	2.556	2.540	2.525	2.509	2.494	2.479	2.464	2.449	2.434	2.420
21	2.405	2.391	2.376	2.362	2.348	2.335	2.321	2.307	2.294	2.280
22	2.267	2.254	2.241	2.228	2.215	2.202	2.189	2.177	2.164	2.152
23	2.140	2.128	2.116	2.104	2.092	2.080	2.068	2.056	2.045	2.034
24	2.022	2.011	2.000	1.989	1.978	1.967	1.956	1.945	1.934	1.924
25	1.913	1.903	1.892	1.882	1.872	1.861	1.851	1.841	1.831	1.821
26	1.812	1.802	1.792	1.782	1.773	1.763	1.754	1.745	1.735	1.726
27	1.717	1.708	1.699	1.690	1.681	1.672	1.663	1.654	1.645	1.637
28	1.628	1.619	1.611	1.602	1.594	1.586	1.577	1.569	1.561	1.553
29	1.545	1.537	1.529	1.521	1.513	1.505	1.497	1.489	1.482	1.474
30	1.466	1.459	1.451	1.444	1.436	1.429	1.421	1.414	1.407	1.400
31	1.392	1.385	1.378	1.371	1.364	1.357	1.350	1.343	1.336	1.329
32	1.323	1.316	1.309	1.302	1.296	1.289	1.282	1.276	1.269	1.263
33	1.256	1.250	1.244	1.237	1.231	1.225	1.218	1.212	1.206	1.200
34	1.194	1.188	1.182	1.176	1.170	1.164	1.158	1.152	1.146	1.140
35	1.134	1.128	1.123	1.117	1.111	1.106	1.100	1.094	1.098	1.083
36	1.078	1.072	1.067	1.061	1.056	1.050	1.045	1.040	1.034	1.029
37	1.024	1.019	1.013	1.008	1.003	0.998	0.993	0.988	0.982	0.977
38	0.972	0.967	0.962	0.958	0.953	0.948	0.943	0.938	0.933	0.928
39	0.924	0.919	0.914	0.909	0.905	0.900	0.895	0.891	0.886	0.881
40	0.877	0.872	0.868	0.863	0.859	0.854	0.850	0.845	0.841	0.837
41	0.832	0.828	0.823	0.819	0.815	0.810	0.806	0.802	0.798	0.794
42	0.789	0.785	0.781	0.777	0.773	0.769	0.765	0.761	0.757	0.753
43	0.749	0.745	0.741	0.737	0.733	0.729	0.725	0.721	0.717	0.713
44	0.709	0.706	0.702	0.698	0.694	0.690	0.687	0.683	0.679	0.676
45	0.672	0.668	0.665	0.661	0.657	0.654	0.650	0.647	0.643	0.640
46	0.636	0.632	0.629	0.625	0.622	0.619	0.615	0.612	0.608	0.605
47	0.602	0.598	0.595	0.591	0.588	0.585	0.582	0.578	0.575	0.572
48	0.569	0.565	0.562	0.559	0.556	0.553	0.549	0.546	0.543	0.540
49	0.537	0.534	0.531	0.528	0.525	0.522	0.518	0.515	0.512	0.509
50	0.506	0.504	0.501	0.498	0.495	0.492	0.489	0.486	0.483	0.480
51	0.477	0.474	0.472	0.469	0.466	0.463	0.460	0.458	0.455	0.452
52	0.449	0.447	0.444	0.441	0.439	0.436	0.433	0.430	0.428	0.425
53	0.423	0.420	0.417	0.415	0.412	0.410	0.407	0.404	0.402	0.399
54	0.397	0.394	0.392	0.389	0.387	0.384	0.382	0.379	0.377	0.375

Table 3 (continued)

θ	0-0	0-1	0-2	0-3	0-4	0-5	0-6	0-7	0-8	0-9
55	0-372	0-370	0-367	0-365	0-363	0-360	0-358	0-356	0-353	0-351
56	0-349	0-346	0-344	0-342	0-339	0-337	0-335	0-333	0-330	0-328
57	0-326	0-324	0-322	0-319	0-317	0-315	0-313	0-311	0-309	0-306
58	0-304	0-302	0-300	0-298	0-296	0-294	0-292	0-290	0-288	0-286
59	0-284	0-282	0-280	0-278	0-276	0-274	0-272	0-270	0-268	0-266
60	0-264	0-262	0-260	0-258	0-256	0-254	0-252	0-250	0-249	0-247
61	0-245	0-243	0-241	0-239	0-237	0-236	0-234	0-232	0-230	0-229
62	0-227	0-225	0-223	0-221	0-220	0-218	0-216	0-215	0-213	0-211
63	0-209	0-208	0-206	0-204	0-203	0-201	0-199	0-198	0-196	0-195
64	0-193	0-191	0-190	0-188	0-187	0-185	0-184	0-182	0-180	0-179
65	0-177	0-176	0-174	0-173	0-171	0-170	0-168	0-167	0-165	0-164
66	0-162	0-161	0-160	0-158	0-157	0-155	0-154	0-152	0-151	0-150
67	0-148	0-147	0-146	0-144	0-143	0-141	0-140	0-139	0-138	0-136
68	0-135	0-134	0-132	0-131	0-130	0-128	0-127	0-126	0-125	0-123
69	0-122	0-121	0-120	0-119	0-117	0-116	0-115	0-114	0-112	0-111
70	0-110	0-109	0-108	0-107	0-106	0-104	0-103	0-102	0-101	0-100
71	0-099	0-098	0-097	0-096	0-095	0-094	0-092	0-091	0-090	0-089
72	0-088	0-087	0-086	0-085	0-084	0-083	0-082	0-081	0-080	0-079
73	0-078	0-077	0-076	0-075	0-075	0-074	0-073	0-072	0-071	0-070
74	0-069	0-068	0-067	0-066	0-065	0-065	0-064	0-063	0-062	0-061
75	0-060	0-059	0-059	0-058	0-057	0-056	0-055	0-055	0-054	0-053
76	0-052	0-052	0-051	0-050	0-049	0-048	0-048	0-047	0-046	0-045
77	0-045	0-044	0-043	0-043	0-042	0-041	0-041	0-040	0-039	0-039
78	0-038	0-037	0-037	0-036	0-035	0-035	0-034	0-034	0-033	0-032
79	0-032	0-031	0-031	0-030	0-029	0-029	0-028	0-028	0-027	0-027
80	0-026	0-026	0-025	0-025	0-024	0-023	0-023	0-023	0-022	0-022
81	0-021	0-021	0-020	0-020	0-019	0-019	0-018	0-018	0-017	0-017
82	0-017	0-016	0-016	0-015	0-015	0-015	0-014	0-014	0-013	0-013
83	0-013	0-012	0-012	0-012	0-011	0-011	0-010	0-010	0-010	0-010
84	0-009	0-009	0-009	0-008	0-008	0-008	0-007	0-007	0-007	0-007
85	0-006	0-006	0-006	0-006	0-005	0-005	0-005	0-005	0-005	0-004
86	0-004	0-004	0-004	0-003	0-003	0-003	0-003	0-003	0-003	0-002
87	0-002	0-002	0-002	0-002	0-002	0-002	0-001	0-001	0-001	0-001
88	0-001	0-001	0-001	0-001	0-001	0-001	0-001	0-000	0-000	0-000

REFERENCES

- BRADLEY, A. J. and JAY, A. H., 1932. *Proc. Phys. Soc.* **44**, 563 ; 1933. *Ibid.* **45**, 507.
 BUEGER, M. J., 1942. *X-ray Crystallography* (New York), p. 414.
 JAY, A. H., 1944. (Forthcoming publication.)
 LIPSON, H. and WILSON, A. J. C., 1941. *J. Sci. Instrum.* **18**, 144.
 TAYLOR, A. and SINCLAIR, H. B., 1945. *Proc. Phys. Soc.* **57**, 126.
 WILSON, A. J. C., 1941. *Proc. Phys. Soc.* **53**, 235.
 WILSON, A. J. C. and LIPSON, H., 1941. *Proc. Phys. Soc.* **53**, 245.

WALL- AND SALT-ABSORPTION CORRECTIONS IN RADIUM-CONTENT MEASUREMENTS

By W. E. PERRY, B.Sc.

Communication from the National Physical Laboratory

MS. received 25 January 1945

ABSTRACT. The measurement of radium content by the gamma-ray method involves corrections for the absorption of the radiation in the wall of the container and in the radioactive material itself. Corrections for cylindrical containers constructed of platinum-iridium, gold, a gold-silver alloy and monel metal have been determined experimentally for wall thicknesses up to 2 mm. and external diameters up to 8 mm, using radium cells and absorbing tubes to simulate radium containers. Measurements of the absorption in powdered materials are described and an empirical formula representing the absorption both for the powdered materials and the metal tubes is deduced. The results provide an experimental basis for the estimation of the absorption in radium salts, and examples of their application to practical cases are given.

§1 INTRODUCTION

THE determination of the radium content of a radium container involves an important correction for the absorption of the gamma rays in the wall of the container and, in the case of certain containers, a correction for the absorption in the radium salt itself. Owen and Naylor (1922) calculated the wall-absorption corrections for platinum, silver and glass containers respectively, using a formula which involves the absorption coefficient of the gamma rays in the material of the container. The corrections for platinum and monel metal were determined experimentally at the National Physical Laboratory by Kaye, Aston and the writer in 1934, nests of absorbing tubes being used as a convenient means of simulating cylindrical radium containers. The measurements covered a range of wall thicknesses from 0.3 to 1.5 mm. and a range of external diameters from 1.5 to 8 mm. It later proved necessary to extend the measurements to wall thicknesses of 2 mm., and the opportunity was therefore taken to repeat the earlier measurements, particularly as radium cells and absorbing tubes became available which were more suitable in several respects than those used previously. These measurements are dealt with in the earlier part of this paper, while the later part deals with measurements undertaken to obtain data on which to base the estimation of the correction for the absorption in the radium salt itself.

§2 ABSORPTION IN CYLINDRICAL METAL TUBES

2.1. *Radium sources and absorbing tubes.** The metals for which absorption measurements have been made are: (a) Platinum-iridium (90% Pt, 10% Ir),

* We are indebted to Messrs Eldorado Gold Mines, Ltd., for the loan of the radium and to Messrs. Johnson, Matthey and Co, Ltd, for the construction and loan of the absorbing tubes used in the measurements.

which will be referred to as platinum; (b) gold; (c) an alloy of gold (67%), silver (20.65%), copper (10.9%) and platinum (1.45%), which will be referred to as gold-silver alloy; (d) monel metal (67% Ni, 28% Cu, 5% Fe, Mn, etc.).

Three radium cells were obtained for the measurements. Each consisted of radium salt enclosed in a cylindrical monel-metal cell, the radium contents being approximately 8, 20 and 20 mg. radium element respectively. The wall thickness of each cell was 0.1 mm., the external length 25 mm. and the external diameters 0.8, 2.0 and 4.0 mm. The cells were closed by solid plugs of monel metal 5 mm. long, so that manipulation with forceps could be carried out without risk of damaging the thin walls. Contact photographs showed that the radium was distributed uniformly throughout the internal length of the cell.

Each radium cell was provided with five closely-fitting platinum absorbing tubes of length 35 mm. and wall thicknesses approximately 0.3, 0.5, 1.0, 1.5 and 2 mm. respectively, i.e. there were in all fifteen platinum tubes. The external diameters of the tubes ranged from 1.4 to 4.8 mm. for the small diameter cell, from 2.6 to 6.0 mm. for the medium diameter cell, and from 4.6 to 8.0 mm. for the largest diameter cell. Two gold tubes of wall thicknesses approximately 1 and 2 mm. were available for each radium cell. In the case of monel metal and the gold-silver alloy, five tubes were made for use with the smallest diameter cell, the wall thicknesses ranging from 0.3 to 2 mm.

*2.2. Measurement of wall thicknesses of absorbing tubes.** The wall thicknesses of the platinum and the gold-silver alloy tubes were determined (1) by measuring the external and internal diameters at the ends with a precision microscope; (2) by calculating the internal diameters, using the known mass, density and external dimensions; (3) by measuring the external diameters with a micrometer screw gauge and the internal diameters by inserting cylinders of known diameters. Method (3) was omitted in the case of the gold and the monel-metal tubes. In general, method (1) gave the smallest values, the results indicating that the walls of the tubes were, on the average, 0.02 mm. thinner at the ends than elsewhere. The values obtained by method (2) were generally intermediate between those by methods (1) and (3), and were accordingly adopted as being the best values of the average wall thicknesses. Incidentally, the density of the platinum-iridium was found by the Metrology Division of the Laboratory to be 21.53, that of the gold 19.21, gold-silver alloy 14.79, and monel metal 8.825, the estimated accuracy being ± 1 part in 1000.

2.3. Absorption measurements. The apparatus used for the absorption measurements consisted of the combined ionization chamber and electroscope (Perry, 1936) used in radium-content tests at the National Physical Laboratory, the thickness of the lead filter on the front of the chamber being 0.5 cm. The measurements consisted of comparing the ionization I_0 produced by a radium cell with the ionization I when the cell was enclosed in a metal absorbing tube, the radium being supported in a wooden stand parallel to the front face of the ionization chamber. The "measured" percentage absorption correction c_m is

* The co-operation of the Metrology Division of the Laboratory is gratefully acknowledged in connection with these measurements.

then given by $c_m = 100 \left(\frac{I_0 - I}{I} \right)$. The absorption corrections were, within the limits of experimental error, independent of the distance between the radium and the chamber for distances normally used, viz., 30 to 120 cm. The majority of the measurements were made at distances between 30 and 60 cm., as these gave convenient ionization currents.

§ 3 RESULTS OF ABSORPTION MEASUREMENTS

3.1. *Wall-absorption corrections* The results of the measurements of the absorption in metal tubes are given in table 1, columns (1) and (2) of which give the external diameters and wall thicknesses, and column (3) the "measured" corrections. As explained in the 1936 paper, owing to the presence of the wall of the radium cell, the measured correction is slightly less than it would be were the absorbing tube completely filled with radium, and an allowance for this has to be made. This allowance was made by multiplying the measured correction c_m by the ratio c_1/c_2 , where c_1 and c_2 are respectively the calculated corrections for the tube completely filled with radium and for the same tube when the cell wall is interposed between the inside of the tube and the radium. An accurate value of the absorption coefficient of the gamma rays in the material of the tube is not necessary for this correction factor, and the values used in the calculations were approximate estimations based on preliminary measurements. The "true" corrections, i.e. the values in column 3 corrected for the presence of the cell wall, are given in column 4.

Table 2 contains the derived values of the corrections for tubes of normal dimensions. In the case of platinum, gold and monel metal, the majority of these values have been obtained by graphical interpolation of the results given in table 1. Graphical extrapolation was necessary to derive the corrections for the thinner-walled tubes of large diameter, but for such tubes the variation of the correction with the external diameter is small, and extrapolation introduces no appreciable uncertainty. In the case of gold-silver alloy, for which the measurements were limited to tubes having diameters between 1.4 and 5 mm., Owen and Naylor's formula (1922) was used as a means of extrapolating the results to obtain the corrections for tubes of larger diameter.

3.2. *Discussion of results.* The values of the corrections given in table 2 represent the percentage by which the observed radium content has to be increased in order to obtain the true content. They are considered to be reliable to within ± 0.3 for the thicker tubes and to within less than this figure for the thinner tubes. It should be noted that the values are only applicable when the measuring apparatus is similar to that used at the National Physical Laboratory, and that modifications in the apparatus, e.g., a change in the lead filter or in the lining of the ionization chamber, introduce variations in the corrections. A slight increase in the corrections is necessary when the distance of the radium from the chamber is less than 30 cm. The increase in the percentage correction in the case of platinum and gold is about 0.1 per mm. of wall thickness when the distance is 20 cm., and 0.05 per mm. of wall thickness when the distance is 25 cm. Incidentally, a change of 0.1 in the absorption

Table 1. Experimental wall-absorption corrections (Wall thickness of monel-metal radium cells 0.1 mm. ; lead filter 0.5 cm. thick.)

External diameter of absorbing tube (mm)	Wall thickness of absorbing tube (mm)	Percentage absorption correction	
		As measured	Corrected for presence of cell wall
<i>Platinum</i>			
1.39 ₇	0.27 ₀	3.0	3.3
1.77 ₈	0.46 ₀	5.5	5.7
2.80 ₉	0.98 ₀	12.2	12.6
3.78 ₇	1.46 ₈	18.5	18.9
4.81 ₃	1.95 ₀	25.0	25.4
2.65 ₀	0.29 ₈	3.8	4.0
3.01 ₅	0.48 ₀	6.1	6.3
4.01 ₃	0.97 ₀	12.7	13.1
5.00 ₉	1.47 ₃	19.4	19.9
6.01 ₀	1.96 ₀	26.1	26.6
4.65 ₆	0.29 ₀	3.6	3.8
4.98 ₆	0.46 ₀	6.2	6.4
5.97 ₇	0.94 ₈	13.1	13.5
7.00 ₃	1.44 ₈	19.9	20.4
7.98 ₇	1.95 ₈	27.3	27.8
<i>Gold</i>			
2.79 ₉	0.94 ₈	10.6	10.9
4.80 ₀	1.93 ₀	22.7	23.1
3.98 ₉	0.96 ₀	11.5	11.9
5.99 ₃	1.95 ₀	23.5	23.9
5.99 ₃	0.94 ₀	11.6	11.9
7.99 ₃	1.94 ₀	24.5	24.9
<i>Gold-silver alloy</i>			
1.45 ₈	0.29 ₈	2.0	2.1
1.85 ₇	0.49 ₈	3.4	3.6
2.83 ₃	0.97 ₈	7.3	7.5
3.84 ₈	1.47 ₃	11.2	11.4
4.90 ₇	2.00 ₀	15.4	15.7
<i>Monel metal</i>			
1.46 ₈	0.30 ₀	0.7	0.7
1.83 ₅	0.48 ₀	1.2	1.2
2.85 ₀	0.99 ₈	2.4	2.5
3.84 ₈	1.44 ₈	3.6	3.7
4.82 ₉	1.93 ₈	4.9	5.0
5.49*	0.33 ₃	0.8	0.8
6.08*	0.59 ₀	1.3	1.3
6.87*	0.97 ₈	2.7	2.9
7.91*	1.47 ₁	4.1	4.3

* Results obtained in a previous investigation (1936), using a glass radium cell of 0.35 mm wall thickness and 4.6 mm external diameter

Table 2. Percentage wall-absorption corrections applicable to the measured radium contents of uniformly filled cylindrical tubes. Lead filter 0.5 cm. thick.

External diameter of tube (mm)	Wall thickness of tube (mm.)									
	0.3	0.4	0.5	0.6	0.8	1.0	1.25	1.5	1.75	2.0
<i>Platinum tubes</i>										
1.5	3.7	4.8	5.9							
2.0	3.9	5.0	6.3	7.4						
2.5	4.0	5.2	6.5	7.7	10.1	12.5				
3.0	4.1	5.3	6.7	7.9	10.4	12.9	15.8			
3.5	4.2	5.5	6.8	8.1	10.7	13.2	16.2	19.0		
4.0	4.3	5.5	6.9	8.2	10.9	13.4	16.6	19.5	22.3	
5.0	4.3	5.7	7.1	8.4	11.1	13.8	17.1	20.2	23.3	26.3
6.0	4.4	5.7	7.2	8.6	11.3	14.0	17.5	20.8	24.0	27.0
7.0	4.4	5.8	7.2	8.7	11.4	14.2	17.8	21.2	24.6	27.7
8.0	4.4	5.8	7.3	8.8	11.5	14.3	18.0	21.6	25.1	28.4
<i>Gold tubes</i>										
1.5	3.4	4.5	5.4							
2.0	3.5	4.7	5.8	6.8						
2.5	3.6	4.8	6.0	7.1	9.2	11.2				
3.0	3.7	4.9	6.2	7.3	9.5	11.6	14.3			
3.5	3.8	5.0	6.2	7.5	9.8	12.0	14.7	17.4		
4.0	3.9	5.0	6.3	7.6	10.0	12.3	15.0	17.8	20.4	
5.0	3.9	5.1	6.4	7.6	10.1	12.7	15.5	18.4	21.2	24.0
6.0	4.0	5.1	6.5	7.6	10.2	12.9	15.8	18.8	21.7	24.7
7.0	4.0	5.1	6.5	7.7	10.2	12.9	16.0	19.1	22.2	25.2
8.0	4.0	5.1	6.5	7.7	10.3	13.0	16.1	19.4	22.6	25.7
<i>Gold-silver alloy tubes</i>										
1.5	2.2	2.8	3.5							
2.0	2.3	3.0	3.7	4.4						
2.5	2.4	3.1	3.8	4.6	6.1					
3.0	2.4	3.1	3.9	4.7	6.3	7.6	9.5			
4.0	2.5	3.2	4.0	4.8	6.3	7.9	9.7	11.7	13.4	
5.0	2.5	3.3	4.0	4.8	6.4	8.0	9.8	11.9	13.8	15.7
6.0	2.5	3.3	4.1	4.9	6.5	8.0	9.9	12.2	14.0	16.0
8.0	2.5	3.3	4.1	4.9	6.5	8.1	10.1	12.4	14.3	16.4
<i>Monel-metal tubes</i>										
1.5	0.7	1.0	1.2							
2.0	0.8	1.1	1.3	1.5						
3.0	0.8	1.1	1.3	1.6	2.1	2.6				
4.0	0.8	1.1	1.4	1.6	2.2	2.8	3.3	3.9	4.5	
5.0	0.8	1.1	1.4	1.6	2.2	2.9	3.4	4.1	4.7	5.2
6.0	0.9	1.1	1.4	1.6	2.2	2.9	3.6	4.2	4.8	5.5
8.0	0.9	1.1	1.4	1.6	2.3	3.0	3.7	4.4	5.0	5.8

correction results in a change of 1 part in 1000 in the derived radium-content value.

The measurements in platinum confirm the earlier observations for wall thicknesses up to 0.8 mm., the differences being not greater than 0.3. For wall thicknesses of 1 mm. and upwards, the new values are slightly less than the earlier values, the differences being less marked the greater the diameter. This discrepancy arises mainly from a slight lack of uniformity of one of the absorbing tubes used in the earlier investigation, in which the wall thicknesses were determined at the ends of the tubes only, by a vernier microscope. These agreed in general with the average values, as calculated from the density, to within 0.02 mm. The wall thickness at the ends of the outer tube of the small diameter "nest" was however 0.05 mm. less than the average thickness. When allowance is made for this, the earlier and present series of measurements agree within the limits of experimental error.

The absorption coefficient of the gamma rays in platinum derived from the results is 1.14 cm^{-1} , i.e. slightly lower than the earlier value, 1.16 cm^{-1} . The coefficients for gold, gold-silver alloy, and monel metal are 1.05 cm^{-1} , 0.70 cm^{-1} and 0.25 cm^{-1} respectively, and the corresponding mass-absorption coefficients are 0.053 for platinum, 0.055 for gold, 0.047 for gold-silver alloy and 0.028 for monel metal. For a lead filter 0.5 cm. thick, these values for platinum and gold are a little greater and that of monel metal a little less than those indicated by the curves which Oddie (1939) obtained from measurements with radium-measuring equipment of somewhat different design.

Table 2 contains sufficient values for the interpolation of corrections for tubes having wall thicknesses between 0.3 and 2 mm. and external diameters between 1.5 and 8 mm. Also, plotting the absorption correction against the absorption coefficient for tubes of identical dimensions yields a series of curves by means of which the corrections for tubes of any material of known absorption coefficient may be obtained. As is shown later, the value of the coefficient can be estimated to an accuracy sufficient for most purposes by the empirical formula (2) in §4.3.

§4. SALT-ABSORPTION CORRECTIONS

Most of the radium containers used in medical practice contain only a few milligrammes of radioactive material, the absorption of the gamma rays in which is negligible. When, however, the quantity of radioactive salt is 20 mg. or more, an appreciable absorption occurs and a "salt-absorption" correction is then necessary. The correction can be calculated (Thirring and Schweidler, 1912; Schweidler, 1912; Paterson *et al.*, 1917; Dufton, 1926; Holuba, 1937; Franz and Weiss, 1940), provided the absorption coefficient of the gamma rays in the salt is known, and a number of investigations of radioactive materials have been made to obtain the information required (Paterson *et al.*, 1917; Owen and Fage, 1921; Behounek, 1934; Franz and Weiss, 1935; Curtis, 1939). Owing to the cost of radium and the experimental difficulties involved, it has not been possible to measure directly the absorption in a radium salt. Attempts have been made at the National Physical Laboratory to measure the absorption by placing a radium container of small dimensions behind a larger one, but these

attempts provided no reliable results because it was not possible to estimate accurately the appropriate wall-absorption correction. It has therefore been necessary to derive the required data indirectly from the results of measurements in which ordinary powdered materials were used as absorbers.

4.1 Absorption measurements in powdered materials. The experimental part of the investigation consisted of measurements of the absorption of the gamma rays in powdered materials the constituents of which ranged from carbon (at. no. = 6) to uranium (at. no. = 92). Measurements were first undertaken in which the powdered material was enclosed in a vessel having plane walls, the radium cell being placed outside the vessel and in contact with the side of the vessel remote from the ionization chamber. These measurements served to indicate the dependence of the absorption on the density and composition of the absorbing specimen, but in order to obtain conditions more nearly similar to those of radium-content determinations, further measurements were made in which the radium cell was surrounded by the absorbing material. Preliminary measurements with vessels having plane walls showed that the absorption under these conditions is slightly less than when the radium is outside the specimen. To approximate more closely to the required conditions it was decided to use cylindrical vessels to contain the powder, and to support the radium cell axially inside the cylinders. Two ebonite cylinders of different diameters were constructed for these measurements. They were of external length 25.5 mm., internal length 20 mm., external diameters 13.4 and 23 mm. and internal diameters 11.3 and 20.4 mm. respectively. The 2-mm. diameter radium cell referred to in § 2.1 was used, and with this cell in position the internal volumes of the cylinders were 1.91 and 6.56 c.c., the absorbing thicknesses of powder when the cylinders were full being 4.65 mm. and 9.20 mm. Each material was finely powdered in an agate mortar before being transferred to the cylinder, and its apparent density was determined from the weight of the powder required to fill the cylinder.

The absorption in a powder was measured by comparing the ionization I_0 when the radium cell was enclosed in the otherwise empty cylinder with the ionization I when the cylinder was full. In deriving the absorption coefficient μ of the gamma rays in the powder from the ratio I/I_0 , account was taken of the curvature of the walls of the radium cell and the cylinder and of the presence of the wall of the radium cell. Assuming a parallel beam of gamma rays, it can be shown that

$$\frac{I}{I_0} = \frac{4}{\pi} \int_0^{\pi/2} \cos^2 \phi \, e^{-\mu[(b^2 - a_1^2 \sin^2 \phi)^{1/2} - (a^2 - a_1^2 \sin^2 \phi)^{1/2}]} d\phi, \quad \dots \dots (1)$$

where a_1 and a are the internal and external radii of the radium source and b is the internal radius of the cylinder containing the powder. Graphs were drawn for each cylinder showing the variation of I/I_0 with μ for values of μ up to 0.3 cm^{-1} , and these graphs were used to obtain the values of μ corresponding to the experimental values of I/I_0 .

4.2 Experimental results. The mean results of the measurements are given in table 3, column 1 of which gives the composition of the material, column 2

the apparent density (g./c.c.), and columns 3 and 4 the measured linear and mass-absorption coefficients μ (cm.⁻¹) and μ/ρ (cm.² g.⁻¹) respectively. The figures in column 5 are referred to later. The accuracy of the values of μ/ρ in column 4 is about $\pm 5\%$, and within these limits μ/ρ was the same for each cylinder, indicating that the absorption coefficient is practically independent of the radial thickness of the absorbing specimen for thicknesses between 0.5 and 1 cm. and of the internal diameter of the containing vessel for diameters between 1 and 2 cm. It is seen that materials containing only elements of low atomic number, i.e. less than that of barium (at. no.=56), have approximately the same mass absorption coefficient, the mean value for these materials being 0.031.

Table 3. Absorption in powdered materials. Radium source on axis of cylindrical vessel containing absorbing powder. Lead filter 0.5 cm. thick.

Absorbing material	Density of absorbing material	Measured linear and mass-absorption coefficient		μ/ρ based on empirical formula $\mu_e = 1.04 \times 10^{-25} + 2.5 \times 10^{-31} Z^3$
		μ	μ/ρ	
	(g./c.c.)	(cm. ⁻¹)	(cm. ² g. ⁻¹)	(cm. ² g. ⁻¹)
U ₃ O ₈	3.95	0.255	0.065	0.065
Th(SO ₄) ₃ · 9H ₂ O	1.37	0.064	0.047	0.047
Bi ₂ (SO ₄) ₃	1.70	0.075	0.044	0.048
PbO	3.92	0.220	0.056	0.056
BaSO ₄	3.05	0.100	0.033	0.034
ZnS	2.42	0.072	0.030	0.031
S	1.31	0.040	0.031	0.032
MgO	0.473	0.015	0.032	0.031
C	0.91	0.029	0.032	0.031
<i>Metal tubes</i>				
Au	19.21	1.05	0.055	0.055
Pt/Ir	21.53	1.14	0.053	0.054
Gold-silver alloy	14.79	0.70	0.047	0.048
Monel metal	8.82 _g	0.25	0.028	0.031

There is, however, a wide range of mass absorption coefficients in the case of those materials containing elements of high atomic number. We have included in table 3 the results of the measurements of the absorption in cylindrical metal tubes referred to in §3.2. It will be noted that the mass-absorption coefficient of monel metal, which is representative of elements of low atomic number, is only slightly less than that of the corresponding powdered materials, while the mass-absorption coefficients of the other metals are intermediate between those of powdered materials containing elements of high atomic number.

4.3. *Derivation of the absorption coefficients of radium salts.* In order to derive the absorption coefficients of radium salts corresponding to the results given in table 3, it was found convenient to deal in terms of the absorption coefficient per electron of each element, $\mu_e = \frac{A}{NZ} \cdot \frac{\mu}{\rho}$, where A and Z are the

atomic weight and atomic number and N is Avogadro's number (6.06×10^{23}). The absorption was considered to consist of two parts, one due to scattering and the other to photo-electric and other processes. It is known that the absorption in light elements is due almost entirely to scattering, the absorption per electron being practically independent of the atomic number. The results in table 3 are in agreement with this, the values of μ_e for these materials being from 0.99×10^{-25} to 1.06×10^{-25} . The contribution of the elements of high atomic number to the total absorption in a material was obtained by first deriving the absorption per molecule $\left(\frac{M}{N} \frac{\mu}{\rho}\right)$, where M is the molecular weight, and subtracting from this the absorption due to the elements of low atomic number, taking μ_e for the latter to be 1.04×10^{-25} . The values of μ_e for the elements of high atomic number so obtained were plotted against the atomic number, and it was found that both for powdered materials and metal tubes the points lay close to the curve* represented by

$$\mu_e = 1.04 \times 10^{-25} + 2.5 \times 10^{-31} Z^3. \quad \dots\dots(2)$$

The mass-absorption coefficients of the materials used in the measurements based on this expression are given in the final column of table 3, and the corresponding mass-absorption coefficients of the most commonly used radium salts are given in table 4.

Table 4. Mass-absorption coefficients of radium salts. Lead filter 0.5 cm. thick. Empirical basis $\mu_e = 1.04 \times 10^{-25} + 2.5 \times 10^{-31} Z^3$.

Radium salt	Mass absorption coefficient μ/ρ (cm ² g. ⁻¹)
RaSO ₄	0.055
RaBr ₂	0.051
RaCl ₂	0.057
RaCO ₃	0.058

§ 5. THEORETICAL SALT-ABSORPTION CORRECTIONS

5.1. *Completely filled container.* When a cylindrical container full of radium salt is situated symmetrically on the axis of the ionization chamber, the axis of the container being perpendicular to that of the chamber, then, assuming a parallel beam of gamma rays, the ratio of the measured ionization I to the ionization I_0 had there been no absorption in the radium salt is given by

$$\frac{I}{I_0} = 1 - \frac{8}{6\pi}(\mu d) + \frac{1}{8}(\mu d)^2 - \frac{4}{45\pi}(\mu d)^3, \quad (3)$$

where d is the internal diameter of the container and μ is the absorption coefficient of the gamma rays in the radium salt (Paterson *et al.*, 1917).

The values of the percentage absorption correction 100 calculated

* See Rutherford *et al.*, 1930 ; Kohlrausch, 1928, and Gray, 1931.

by means of (3) for values of μd up to 0.20 are given in table 5. It is seen that the percentage correction is approximately equal to $43\mu d$.

Table 5. Theoretical salt-absorption corrections (%).

μd	0.01	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.20
Percentage absorption correction	0.4	0.9	1.7	2.6	3.4	4.3	5.2	6.0	6.9	7.8	8.7

5.2. *Partially filled container.* Franz and Weiss (1940) have shown that when a cylindrical container is only partially filled with radioactive material the absorption in the material is given by

$$\frac{I}{I_0} = 1 - a(\mu d) + b(\mu d)^2, \quad (4)$$

in which the values of a and b depend on the ratio h/d , h being the depth of the radium salt when the axis of the container is horizontal. The ratio of the correction for a partially filled container to that for a completely filled container is nearly independent of μd for values of μd up to 0.2. Values of the ratio for different values of h/d are given in table 6. It is seen that the correction for a container less than half full is smaller than that for a full container, while the correction for a container more than half but not completely full slightly exceeds that for a full container. The formula (4) is based on the assumption that radon remains occluded in the radioactive material, which is true in the case of a number of radium salts, e.g. radium sulphate, crystalline radium chloride and radium bromide.

Table 6. Ratio of salt-absorption corrections for full and partially full containers

h/d	0.15	0.25	0.5	0.75	0.85	1.0
$\frac{\text{"Partially filled" correction}}{\text{"Completely filled" correction}}$	0.64	0.8	1.0	1.03	1.04	1.0

§ 6 PRACTICAL SALT-ABSORPTION CORRECTIONS

6.1. *Concentrated radium salts.* Tables 4, 5 and 6 contain the necessary data for the estimation of salt-absorption corrections when the type and density of the salt and the internal diameter of the container are known. Consider, for example, a cylindrical tube of internal length 8 mm. and internal diameter 4 mm. containing 200 mg. Ra element in the form of radium sulphate which completely fills the tube. The weight of the salt is 285 mg., the density (ρ) is 2.84 g./c.c., the mass-absorption coefficient (μ/ρ) 0.055 and the linear absorption coefficient

μ is 0.156 cm^{-1} . The value of μd ($d = 4 \text{ mm.}$) is thus 0.0624 and the salt-absorption correction interpolated from table 5 is 2.7% . When the corrections are of this order of magnitude, a 5% change in the value of μ changes the correction by 0.1 and the radium-content value by 1 part in 1000 . Table 7 gives the corrections applicable to full tubes of radium sulphate for packing densities of $2, 2.5$ and 3 g./c.c. The corresponding corrections for tubes of radium bromide are a little less and those for radium chloride and radium carbonate a little greater than the corrections given in table 7.

Table 7. Percentage salt-absorption corrections for full tubes of radium sulphate. Lead filter 0.5 cm. thick.

Internal diameter of tube (mm.)	Percentage correction		
	$\rho = 2 \text{ g./c.c.}$	$\rho = 2.5 \text{ g./c.c.}$	$\rho = 3 \text{ g./c.c.}$
1.0	0.5	0.6	0.7
3.0	1.4	1.8	2.1
6.0	2.8	3.6	4.2
8.0	3.8	4.7	5.7

6.2. Radioactive luminous compounds. Radioactive luminous compound consists of a mixture of zinc sulphide with radium or with radium and mesothorium, the radioactive material being present in concentrations of the order of a hundred microgrammes of radium element or its equivalent per gramme of compound. Paterson, Walsh and Higgins (1917) determined the absorption in zinc sulphide from the results of gamma-ray measurements before and after mixing a known quantity of radium salt with zinc sulphide and also by surrounding a sealed tube of radium salt with a known thickness of zinc sulphide. By the former method the absorption coefficient was found to be 0.3 cm^{-1} and by the latter 0.092_5 cm^{-1} . It was later (1921) suggested by Owen and Fage that the higher value could have been due to the imperfect sealing of the tube containing the mixture of radium salt and zinc sulphide. These authors found the absorption coefficient to be 0.101 cm^{-1} , the value being the same, within the limits of experimental error, for zinc sulphide mixed with radium as for zinc sulphide alone. The mass-absorption coefficient corresponding to the results of Owen and Fage is 0.050 , i.e. greater than the present author's value, 0.030 . It seems that different scattering conditions* may account at least in part for the divergent values. A contributory cause may be 'alterations in the materials and disposition of the measuring apparatus which have been introduced since the earlier measurements were made.

The corrections for full tubes of luminous compound 3 cm. or less in diameter,

* A value approaching 0.050 was obtained by the author when the cross-section of the absorbing specimen (zinc sulphide enclosed in a flat-walled vessel) was only just sufficient to screen the ionization chamber from the radium cell, the latter being in the "end-on" position with respect to the chamber. Correspondingly high values were also obtained in the case of other powdered materials. This arrangement of source and absorber decreases the amount of radiation scattered into the ionization chamber by the absorbing specimen, partly on account of the reduced scattering area and partly on account of the partial separation of the source and the specimen (Tarrant, 1931/32.)

based on the mass-absorption coefficient 0.03 for zinc sulphide, are given sufficiently accurately by $c_s = 1.3\rho d$, where c_s is the percentage correction, ρ is the density of the compound in g./c.c. and d is the internal diameter of the tube in cm. The additional correction for the absorption in the walls of the glass tube in which such compound is usually enclosed for measurement purposes is 0.8% per mm. of wall thickness for thicknesses up to 2 mm. The same corrections apply when mesothorium is present in addition to radium, since the absorption of the gamma rays from mesothorium and its decay products is only slightly different from that of the gamma rays from radium.

6.3. *Radium concentrates.* "Radium concentrates" is the term generally applied to radium-bearing products of refining operations. Such products usually contain several milligrammes of radium element per kilogramme of material, the latter comprising a high proportion of barium and frequently some lead, the remaining constituents being elements of low atomic number. A convenient sample for radium-content tests is one hundred grammes accommodated in a sealed glass cylinder about 5 cm. in diameter and 10 cm. long. The mass-absorption coefficient of the material depends on its composition, and is derived by using the values given in table 3, or by using the empirical formula (2) for the various constituents of the material. A number of such samples sent recently to the Laboratory for radium-content tests contained 35% to 45% barium oxide (BaO) and 5% to 15% lead oxide (PbO), the remainder consisting of elements of low atomic number. The mass-absorption coefficients of the samples derived by using (2) ranged from 0.034 to 0.038 and the salt-absorption corrections from 6% to 8%.

6.4. *Radium-beryllium neutron sources.* Radium-beryllium neutron sources consist of an intimate mixture of radium salt with beryllium powder in the proportion of about 100 mg. of radium element to several hundred mg. of beryllium. The mass-absorption coefficient of beryllium may be assumed to be the same as that for the elements of low atomic number in table 3, i.e. 0.032. Curtis (1939) has determined experimentally the total wall- and salt-absorption correction for a neutron source consisting of 1237 mg. of beryllium and 139.7 mg. of radium element (in the form of radium sulphate) enclosed in a brass cylinder of internal diameter 8 mm., internal length 24.5 mm. and wall thickness 1 mm. Assuming the brass to be of density 8.55, the wall-absorption correction, based on the results given in this paper, is 3.0%. The salt-absorption correction is 1.4%, the total correction being thus 4.4%. Curtis obtained the value 5.1% by radium-content measurements before and after mixing the radium salt and the beryllium. The difference between these two values is probably due in part to differences in the measuring equipment in the two cases.

ACKNOWLEDGMENTS

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory. The author desires to acknowledge the assistance with the measurements rendered by past and present members of the Radiology Section of the Physics Division of the Laboratory.

REFERENCES

- BEHOUNEK, F., 1934. *Z Phys.* **92**, 533.
 CURTIS, L. F., 1939. *Bur. Stand. J. Res., Wash* **23**, 617.
 DUFTON, A. F., 1926 *Brit. J. Radiol* **22**, 89, 157
 FRANZ, H. and WEISS, C., 1935. *Phys. Z.* **36**, 486, 1940. **41**, 345.
 GRAY, L. H., 1931 *Proc Camb. Phil Soc* **27**, 103.
 HOLUBA, M., 1937. *Sitzber Akad Wiss Wien*, **11** (a), 146, 285.
 KAYE, G. W. C., ASTON, G. H. and PERRY, W. E., 1934 *Brit. J. Radiol.* **7**, 540.
 KOHLRAUSCH, K. W. F., 1928 *Handb. Experimental-Physik*, **15**, 101.
 ODDIE, T. H., 1939. *Proc. Phys. Soc.* **51**, 905
 OWEN, E. A. and FAGE, W. E., 1921. *Proc. Phys. Soc.* **34**, 27.
 OWEN, E. A. and NAYLOR, B., 1922. *Proc. Phys. Soc.* **34**, 92
 PATERSON, C. C., WALSH, J. W. T. and HIGGINS, W. F., 1917. *Proc. Phys. Soc.* **29**, 215
 PERRY, W. E., 1936 *Brit J Radiol.* **9**, 648.
 RUTHERFORD, E., CHADWICK, J. and ELLIS, C. D., 1930. "Radiations from Radioactive Substances", p 479
 SCHWEIDLER, E., 1912. *Phys. Z.* **13**, 453.
 TARRANT, G. T. P., 1931-32. *Proc. Camb. Phil. Soc* **28**, 475.
 THIRRING, H. and v. SCHWEIDLER, E., 1912. *Phys. Z.* **13**, 266.

THEORETICAL SHAPE OF THE COMPTON PROFILE FOR ATOMS FROM H TO Ne

BY W. E. DUNCANSON AND C. A. COULSON,
 University College, London, and University College, Dundee

MS received 20 December 1944

ABSTRACT. Closed analytical expressions are obtained for the shapes of the modified Compton lines scattered from the elements H to Ne, in atomic form. The widths of the Compton lines are calculated for these atoms, and conditions of resolution from the unmodified lines are obtained. Certain errors in other work on this subject are noted and corrected. The width of the Compton line, as determined experimentally, for diatomic molecules is about 15 to 25 % greater than for isolated monatomic atoms.

§ 1. INTRODUCTION

IT is well known that a study of x-ray scattering enables us to determine the spatial distribution of electrons within an atom: in this way the wave-mechanical predictions have been fully confirmed. It is, however, also possible to determine the velocity distribution, and again this may be compared with the theoretical predictions. This is achieved by a study of the shape of the Compton modified line. For if a homogeneous beam of x rays falls on an atom in which all the electrons are at rest, the Compton line should have a single modified frequency. But if the electrons are moving there is a Doppler effect, related to the velocities of the electrons; this causes the modified line to spread out and become a band, whose profile determines the velocity distribution. This is the photon theory of Jauncey and DuMond (DuMond, 1933), which has been shown by Burkhardt (1936) and Schnaidt (1934) to give effectively the same results as the complete wave-mechanical theory due to Wentzel and Waller.

Recent improvements in x-ray technique (Kappeler, 1936) enable an accurate determination of the Compton profile to be made. We therefore present in this paper the theoretical profiles to be expected for the atoms H to Ne. The general appearance of these curves has, indeed, already been calculated by Kirkpatrick, Ross and Rutland (1936, referred to hereafter as K.R.R.). These calculations, however, are not satisfactory; there are three reasons for this. In the first place relatively inaccurate wave-functions were used, in which there was not sufficient flexibility; in the second place it was assumed that the individual atomic orbitals were mutually orthogonal, when in fact (see later) they are not; and finally there appear to be discrepancies in the actual calculated values, whose origin we have not been able to discover. Thus Hicks (1937) has calculated the Compton profile for He using various approximate wave-functions. We have performed some of his calculations independently, and agree with him; but we disagree with the values of K.R.R., which are quoted by Hughes and Mann (1938). An even more serious disagreement occurs for atomic H, where the values (quoted by Hughes and Starr, 1938, who were equally puzzled by the discrepancy) given by Hicks are reproduced in our own work; they differ from the K.R.R. values by as much as 25% in the important part of the curve. We find a similar disagreement with the carbon curves, though we are unable to explain why the K.R.R. values are incorrect. In view of these differences we have thought it worth while to recalculate the shape of the Compton profile for all the atoms H to Ne, taking advantage of the fact that more accurate wave-functions are now available (Duncanson and Coulson, 1944).

These calculations have another importance, for Hughes and Mann (1938) have shown that the distribution of inelastically scattered electrons has the same form as the function describing the Compton profile. Indeed, one reason for the present calculations is that accurate experiments of this kind have been made for H₂ and He, and their conclusions agree very precisely with those predicted theoretically.

§ 2. METHOD OF CALCULATION

Our method of calculation follows closely that which has been used in an earlier series of papers by the present authors (for a complete list of references see Duncanson, 1943). Thus let p be the momentum of an electron in the atom, and $I(p)$ the mean radial distribution function. This means that $I(p)dp$ is the probability that p has a magnitude between p and $p + dp$. Obviously

$$\int_0^{\infty} I(p)dp = 1. \quad \text{.....(1)}$$

$I(p)$ may be calculated at once if we know the momentum wave-function $\chi(p)$. Thus, if there is only one electron,

$$I(p) = \int \chi(p)\chi^*(p)p^2 d\omega, \quad \text{.....(2)}$$

where $d\omega$ is an element of solid angle for p . The extension required for more than one electron is obvious. Now the momentum wave-functions $\chi(p)$ are not known directly, but may be obtained from the space wave-functions $\psi(r)$

by the Dirac transformation theory. Thus, for one electron, using atomic units,

$$\chi(p) = (2\pi)^{-3/2} \int e^{-ip \cdot r} \psi(r) dr. \quad \dots\dots (3)$$

Again, as in (2), the extension to more electrons is straightforward.

The actual shape of the Compton profile depends on the incident wave-length λ_i and the angle of scattering θ . But all angles and wave-lengths are included in one single formula in the following way. Let λ_c be the peak of the modified line, so that the familiar relation holds:

$$\lambda_c = \lambda_i + 2\gamma \sin^2 \theta/2, \quad \dots\dots (4)$$

where

$$\gamma = h/m_0c = 0.0243 \text{ A.}$$

Then choose a new wave-length λ^* defined by

$$2\lambda^* = (\lambda_c^2 + \lambda_i^2 - 2\lambda_c \lambda_i \cos \theta)^{1/2}. \quad \dots\dots (5)$$

In most cases effectively

$$\lambda^* = \lambda_i \sin \theta/2. \quad \dots\dots (6)$$

Next let J be the intensity of the Compton band at a displacement l from λ_c , and let us introduce the variable

$$q = cl/2\lambda^*, \quad \dots\dots (7)$$

where c is the velocity of light (value 137 in atomic units). Then J is a function merely of q , with its peak at $q=0$, symmetrical about the peak, and defined by

$$J(q) = \frac{1}{2} \int_q^\infty \frac{I(p) dp}{p}. \quad \dots\dots (8)$$

Previous workers have always used an arbitrary constant k instead of $\frac{1}{2}$ in this formula, but we can show that $\frac{1}{2}$ is necessary if we want the total area below the $J(q)$ curve to be unity. In that case we can speak of a normalized profile. For using (8) the area under the $J(q)$ curve is

$$\begin{aligned} \int_{-\infty}^{+\infty} J(q) dq &= 2 \int_0^\infty J(q) dq = \int_0^\infty \int_q^\infty \frac{I(p) dp}{p} dq \\ &= \int_0^\infty \int_0^p \frac{I(p)}{p} dq dp \\ &= \int_0^\infty I(p) dp \\ &= 1, \text{ from (1).} \end{aligned}$$

Equation (8) enables us to calculate the Compton profile, since (2) and (3) have already provided us with a knowledge of $I(p)$.

§ 3. WAVE-FUNCTIONS USED

The best available wave-functions for the series of atoms concerned are those recently described by Duncanson and Coulson (1944). They are formed by suitable linear sums of determinants compounded of the separate orbitals:

$$\begin{aligned} \psi(1s) &= \sqrt{(\alpha^3/\pi)} e^{-\alpha r}, \\ \psi(2s) &= \sqrt{(\mu^5/3\pi N)} [re^{-\mu r} - (3A/\mu) e^{-\beta r}], \\ \psi(2p_0) &= \sqrt{(\gamma^5/\pi)} r \cos \theta e^{-\gamma r}, \\ \psi(2p_\pm) &= \sqrt{(\gamma^5/2\pi)} r \sin \theta e^{-\gamma r \pm i\phi}. \end{aligned} \quad \dots\dots (9)$$

Values of the parameters $\alpha(\equiv\mu a)$, $\beta(\equiv\mu b)$, $\gamma(\equiv\mu c)$, μ , A and N are given for each separate atom. The states of the atoms with which we are concerned are

H	$(1s)^2 S$,	He	$(1s)^2 {}^1S$,	Li	$(1s)^2(2s)^2 S$,
Be	$(1s)^2(2s)^2 {}^1S$,	B	$(1s)^2(2s)^2(2p)^2 P$,	C	$(1s)^2(2s)^2(2p)^2 {}^3P$,
N	$(1s)^2(2s)^2(2p)^3 {}^4S$,	O	$(1s)^2(2s)^2(2p)^4 {}^3P$,	F	$(1s)^2(2s)^2(2p)^5 {}^2P$,
Ne	$(1s)^2(2s)^2(2p)^6 {}^1S$.				

Other states, and particularly solid or crystal states, would require different wave-functions, and the Compton profiles would not be given by our present calculations. But for atomic states involving the same numbers of s and p orbits, differently arranged, it appears that the profiles would be almost unaffected. Previous to the availability of (9) it had been common to use Slater wave-functions (1930), and we have therefore carried through some of our calculations for both types for comparison. Slater's wave-functions are particular cases of (9) with $A=b=0$, $\mu=\gamma$, $N=1$, and with different values of the exponents α and μ .

§ 4 FORMULAE

If all the $1s$, $2s$, $2p$ orbitals are mutually orthogonal, it is not hard to show (Duncanson and Coulson, 1941) that $I(p)$ is simply the sum of a suitable number of separate functions $I_{1s}(p)$, $I_{2s}(p)$, $I_{2p}(p)$, and consequently, from (8), that $J(q)$ is itself the sum of a corresponding number of functions $J_{1s}(q)$, $J_{2s}(q)$, $J_{2p}(q)$. This illustrates a contention of Hicks (1940) that the over-all line shape for the whole atom is the sum of the component line shapes of each type of electron.

We omit the details of the calculations of the integrals, but it follows from (2), (3), (8) and (9) that

$$\begin{aligned}
 I_{1s}(p) &= 32\alpha^5 p^2 / \pi(p^2 + \alpha^2)^4, \\
 I_{2s}(p) &= \frac{32\mu^5 p^2}{3N\pi} \left\{ \frac{3\mu^2 - p^2}{(p^2 + \mu^2)^3} - \frac{3A}{\mu} \frac{\beta}{(p^2 + \beta^2)^2} \right\}^2, \\
 I_{2p}(p) &= 512\gamma^7 p^4 / 3\pi(p^2 + \gamma^2)^6, \\
 J_{1s}(q) &= 8\alpha^5 / 3\pi(q^2 + \alpha^2)^3, \\
 J_{2s}(q) &= \frac{16\mu^5}{3N\pi} \left\{ \frac{5q^4 - 20q^2\mu^2 + 23\mu^4}{30(q^2 + \mu^2)^5} + \frac{3\beta^2 A^2}{2\mu^2(q^2 + \beta^2)^3} \right. \\
 &\quad \left. - \frac{3\beta A}{\mu(\mu^2 - \beta^2)^4} \left[2(5\mu^2 + \beta^2) \log \frac{q^2 + \beta^2}{q^2 + \mu^2} + \frac{2\mu^2(\mu^2 - \beta^2)^2}{(q^2 + \mu^2)^2} \right. \right. \\
 &\quad \left. \left. + (\mu^2 - \beta^2) \left(\frac{3\mu^2 + \beta^2}{q^2 + \beta^2} + \frac{7\mu^2 + \beta^2}{q^2 + \mu^2} \right) \right] \right\}, \\
 J_{2p}(q) &= \frac{32\gamma^7}{15\pi} \frac{5q^2 + \gamma^2}{(q^2 + \gamma^2)^5}.
 \end{aligned}
 \tag{10}$$

In this way a closed analytical expression is obtained for the shape of the Compton line.

If the component orbitals are not orthogonal, as occurs with the Slater wave-functions, neither $I(p)$ nor $J(q)$ is of the simple form represented by sums of contributions from each electron, and new terms appear, which we do not need to describe here, due to the non-orthogonality of $\psi(1s)$ and $\psi(2s)$. In this case

the calculations may be much simplified by defining a new $2s$ orbital of the form $\psi(2s) + m\psi(1s)$, and choosing m so that this is orthogonal to $\psi(1s)$. We may describe this as orthogonalizing the Slater wave-functions (Slater, 1932; Wilson, 1933).

It may perhaps be mentioned here that the wave-functions used by K.R.R. were hydrogen-like. That is to say, the $2s$ and $2p$ orbitals, like the $1s$ orbital, were those to be expected for a single electron moving round an arbitrary central charge, whose value was found by using the rules (but not the wave functions) developed by Slater. These orbitals are only orthogonal if the same effective nuclear charge is used for all electrons, except that if their l and m quantum numbers differ, they are orthogonal by symmetry. With the values used by K.R.R. and, later, by Burkhardt (1936), a different effective charge is used for the K and L shell electrons. One possible, though, as we believe, small, source of error arises in this respect.

§ 5. RESULTS AND DISCUSSION

We show in figure 1 the normalized $J(q)$ curves for H, He, Li, C and Ne, and in figure 2 the curves for Li to Ne adjusted so that the peak value is unity in each case. This enables a better comparison to be made. Only half the profile is shown since, to our present approximation, the complete profile is symmetrical. The corrections to be made for relativity and binding are small, and Burkhardt

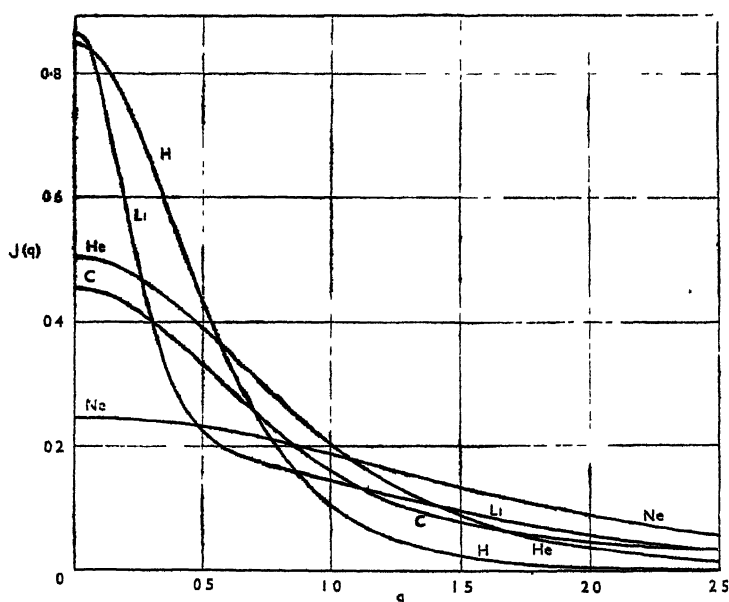


Figure 1

(1936) has shown that they exert only a small effect upon the symmetry of the profile. The general features of these curves are similar to those described by K.R.R. : that is to say, s electrons give sharp profiles and p electrons wider ones, and as we proceed along a row of the periodic table, the profile becomes steadily broader, narrowing again for the alkali metal that starts the new sequence.

The shapes of the curves are conveniently summarized by the total width at half-maximum. The corresponding q -values, written Δq , are given in the first row of table 1. The width may be expressed on a wave-length scale by recalling (equations (6) and (7)) that $q = cl/2\lambda_e \sin \theta/2$, and that the width at half-maximum is Δl , where $\Delta l = 2(\lambda_e/c) \sin \theta/2 \Delta q$. In the bottom row of the

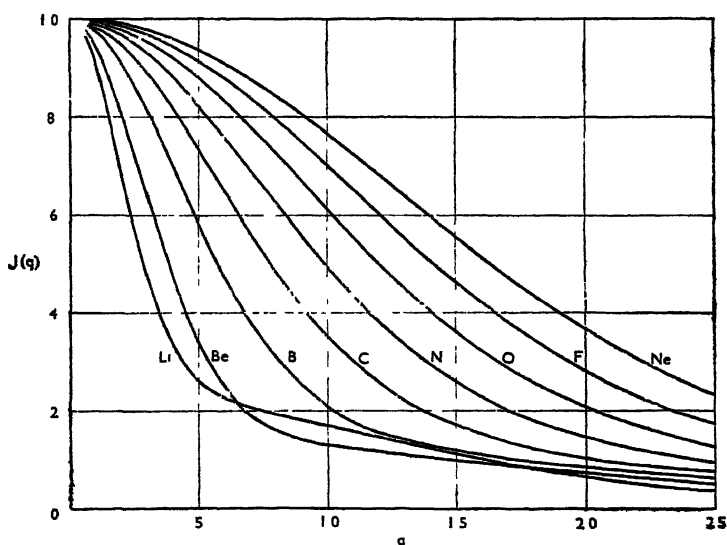


Figure 2.

table we give the corresponding values of Δl for $\theta = 180^\circ$ and $\lambda_e = 7076$ x.u. (Mo $K\alpha$ radiation.) The gradual increase of Δl as we go along the periodic table from Li to Ne is clearly shown. Wollan's contention (1934) that all atoms from $Z=3$ to $Z=13$ should have $\Delta l = 22$ x.u. is evidently not correct, even though allowance be made for possible inadequacies in our theoretical wave-functions.

Table 1. Half-widths of Compton profiles

	H	He	Li	Be	B	C	N	O	F	Ne
Δq	1.02	1.74	0.57	0.77	1.15	1.57	1.98	2.40	2.84	3.27
Δl (x u.) Mo $K\alpha$ $\theta = 180^\circ$	11	18	5.9	8.0	12	16	20	25	29	34

Let us consider what experiments are available with which to check the Δl and Δq values of table 1. In the first place the calculated values are for monatomic systems, this limits a strict comparison to He and Ne. For neon (Burkhardt, 1936; Kappeler, 1936) the experimental value is $\Delta l = 32$ x.u., so that $\Delta q = 3.1$, in excellent agreement with theory. For He, DuMond and Kirkpatrick (1937) obtain an experimental value $\Delta l = 16$ x.u., giving $\Delta q = 1.5$. This, indeed, was the value deduced theoretically by Hicks (1937) using much more involved wave-functions. As there are only two electrons in He, it is possible to refine the wave-functions considerably by explicit inclusion of the distance between the electrons. Hicks shows that this reduces Δq from about 1.7 to 1.5. As the mutual

electron repulsion is relatively more important in He than in the other atoms concerned, we may reasonably expect a smaller correction term in the rest of the table. It may be mentioned here also that Hughes and Mann (1938), in electron-scattering experiments, obtain effectively the same Δq for He as DuMond and Kirkpatrick in their x-ray experiments.

It thus appears that our curves fit satisfactorily with experiment for He and Ne. There appear to be no other experiments directly suitable for comparison. The pioneer work of Woo (1926) does indeed give shapes for these atoms, and in general terms they agree with figure 2. But as K.R.R. say, Woo "sacrificed resolving power in favour of intensity", and as a result "all details of the shape were missed". There is a further difficulty in that Woo necessarily used solid scatterers, whereas our formulae relate to free atoms. In a solid, as the experiments of DuMond with graphite show quite clearly (DuMond, 1933, figure 33), the velocities of the electrons are greater than in the free atom, and the Compton profile is correspondingly broadened.

It is however possible to compare our table of Δq values with the results of experiments on diatomic molecules. The effect of molecular binding is to increase the mean momentum of the electrons, and hence to broaden the Compton profile. Thus Duncanson and Coulson (1941) and Coulson and Duncanson (1942) have shown that the CH_4 profile is about 5% wider than that corresponding to $\text{C} + 4\text{H}$; and Duncanson (1943) shows that much the same occurs for N_2 and Li_2 , the increase still being of this order. Experimentally, Kappeler (1936) gives $\Delta q = 2.7$ for O_2 and $\Delta q = 2.5$ for N_2 , where we should have expected 2.4 and 2.0 respectively for free atoms. The enhanced value of Δq is clearly shown, though its size is rather more than we should have anticipated. The increase is, however, actually as large as claimed, for Hughes and Starr (1939) get just the same value for N_2 with electron scattering. In the case of H_2 , Hicks (1937) computes $\Delta q = 1.17$, about 14% greater than the atomic value, and Hughes and Starr (1938) get an experimental value about 11% wider than this. It would be very valuable to have accurate x-ray values for H_2 . At present it appears that by forming a diatomic molecule, the width of the Compton profile is increased by about 15 to 25%.

The only other reasonably valid experimental result which we can utilize is Kappeler's value (1936) for acetylene carbon black. This gives $\Delta q = 2.0$, again about 25% greater than the atomic value. This type of carbon is probably more nearly molecular than crystalline, so that the value is about what might have been expected.

If we consider pure crystals, the width is much greater than for free atoms. Thus with graphite there is about 75% excess, and with Be 140% excess above the atomic values. It is evidently unreasonable to attempt any comparison; the profile needs to be calculated entirely differently, as was done for C and Be by Bloch (1934). The situation is even worse with metallic Li (Kappeler, 1936).

With longer incident waves the Compton profile broadens and finally merges into the unmodified line. The curves in figure 2 may be used to indicate the conditions under which the two lines may be resolved. It is well known (Wollan, 1933) that for light elements the modified line is stronger than the unmodified,

and that for heavy elements the unmodified line is the stronger. No completely rigorous discussion of resolving the two can be given without taking this variation into account, but a rough criterion can be made by supposing that if at the point midway between the maxima of the two bands the intensity of the Compton band is less than one-half of its maximum value, then resolution is possible. This convention, which is almost certainly too severe, probably by a factor of nearly two, will at least indicate the order of wave-length necessary for resolution, and show how it varies with the atom concerned.

Now the position of the unmodified line ($\lambda = \lambda_0$) is given on the q -scale by

$$q = \frac{c(\lambda_c - \lambda_0)}{2\lambda_0 \sin \theta/2} = \frac{c\gamma \sin \theta/2}{\lambda_0} \text{ approx.} \quad \dots\dots(11)$$

So at the mid-point $q = \frac{c\gamma \sin \theta/2}{2\lambda_0}$, and our criterion requires that for this value

of q , $J(q) < \frac{1}{2}J(0)$. This gives upper limits for $\frac{\lambda_0}{\sin \theta/2}$, which are shown (in A.) in table 2. It appears that large angles of scattering are less favourable for this resolution, but the effect is not great unless θ is small, and then all the displacements are so small as to decrease the accuracy of measurement on account of scatter. This table also shows how the resolution decreases for heavier atoms along each row of the periodic table.

Table 2. Maximum values (in A.) of $\frac{\lambda_0}{\sin \theta/2}$ for resolution of modified and unmodified lines

H	He	Li	Be	B	C	N	O	F	Ne
3.3	1.9	5.9	4.2	2.9	2.1	1.7	1.4	1.2	1.0

In conclusion, we may refer briefly to two other related matters. In the first place, as figures 1 and 2 show, the shape of the Compton profile varies from atom to atom. This variation is not only in width, but also particularly in the relative importance of the "leg" at the base. This means that the approximate shape $y = a / \left\{ 1 + \left(\frac{x}{b} \right)^2 \right\}$ assumed by Hoyt (1932), and employed subsequently by other workers, is only a fair representation over the greater part of the band for atoms where the effect of the s electrons in sharpening the profile is counter-balanced by a larger number of p electrons broadening it. In the cases of Li and Be in particular, and presumably also for other elements in the same columns of the periodic table, the presence of the tail will complicate the separation of the contributions to the full profile that arise from satellite lines, as, e.g., $K\alpha_1$ and $K\alpha_2$ doublet, if Hoyt's formula is used. The method used by DuMond (1933, p. 31) appears more suitable.

Our final comment concerns the use of orthogonal orbitals, which we referred to at the end of § 4. Since the $2p$ orbital is by symmetry orthogonal to the $1s$ and $2s$ orbitals, we are here only concerned with the lack of orthogonality between the $1s$ and $2s$ wave-functions. We have made detailed calculations of this effect for one or two atoms. For convenience we used the somewhat simpler wave-functions of Slater rather than the more elaborate ones used in the earlier

part of this work. This should not, however, greatly affect the error due to neglect of non-orthogonality. It appears from our work that an alteration not exceeding 5% is made in any one value of J , the result of neglecting non-orthogonality is to widen the theoretical profile very slightly. Table 3 shows for the atoms Li and C the widths Δq at half-maximum value when (a) the non-orthogonality is neglected and (b) when the $2s$ orbital is suitably orthogonalized. The effect on the width is surprisingly small, and suggests that it should be possible to calculate sufficiently accurate profiles for other elements in the same simple way, without recourse to the labour of orthogonalization.

Table 3. Effect of non-orthogonality

Atom	Δq , neglecting non-orthogonality	Δq , allowing for non-orthogonality
Li	0.571	0.567
C	1.574	1.562

REFERENCES

- BLOCH, 1934. *Phys. Rev.* **46**, 674.
 BURKHARDT, 1936. *Ann. Phys., Lpz.*, **26**, 567.
 COULSON and DUNCANSON, 1942. *Proc. Camb. Phil. Soc.* **38**, 100.
 DUMOND, 1933. *Rev. Mod. Phys.* **5**, 1.
 DUMOND and KIRKPATRICK, 1937. *Phys. Rev.* **52**, 419.
 DUNCANSON, 1943. *Proc. Camb. Phil. Soc.* **39**, 180.
 DUNCANSON and COULSON, 1941. *Proc. Camb. Phil. Soc.* **37**, 406.
 DUNCANSON and COULSON, 1944. *Proc. Roy. Soc. Edinb.* **62 A**, 37.
 HICKS, 1937. *Phys. Rev.* **52**, 436.
 HICKS, 1940. *Phys. Rev.* **57**, 665.
 HOYT, 1932. *Phys. Rev.* **40**, 477.
 HUGHES and MANN, 1938. *Phys. Rev.* **53**, 50.
 HUGHES and STARR, 1938. *Phys. Rev.* **54**, 189.
 HUGHES and STARR, 1939. *Phys. Rev.* **55**, 343.
 KAPPELER, 1936. *Ann. Phys., Lpz.*, **27**, 129.
 KIRKPATRICK, ROSS and RITLAND, 1936. *Phys. Rev.* **50**, 928.
 SCHNAIDT, 1934. *Ann. Phys., Lpz.*, **21**, 89.
 SLATER, 1930. *Phys. Rev.* **36**, 57.
 SLATER, 1932. *Phys. Rev.* **42**, 33.
 WILSON, 1933. *J. Chem. Phys.* **1**, 210.
 WOLLAN, 1933. *Phys. Rev.* **43**, 955.
 WOLLAN, 1934. *Phys. Z.* **35**, 353.
 WOO, 1926. *Phys. Rev.* **27**, 119.

ACHROMATIZED PLATE-MIRROR SYSTEMS

By E. H. LINFOOT,

University of Bristol

MS. received 11 December 1944

ABSTRACT. A variant of the Schmidt-Cassegrain optical system is described in which the single aspheric plate is replaced by two plates, of different glasses and opposite asphericities, whose colour-errors compensate each other. The use of two plates makes it possible to obtain not only apochromatism but also flat-fielded anastigmatism with both mirrors spherical. Design data are given for such two-sphere two-plate anastigmats and their distortion is shown to be small. Next, the effect of relaxing the strict anastigmatism condition is considered and it is shown that the admission of a small amount of Seidel astigmatism leads to a large extension of the range of available systems. Finally two aplanat types are selected which seem well suited to astronomical application.

§ 1 INTRODUCTION

THE colour-error in a Schmidt-Cassegrain camera, chiefly chromatic variation of spherical aberration, is small enough to be tolerated in systems of moderate size. For example, in an $F/2$ system of focal length 13 inches, constructed with a borosilicate crown plate, the diameters of the red and blue confusion circles on axis are each about 0.0015 inch when the plate is figured in green mercury light, while in an $F/2.5$ system of the same focal length their size is only half this amount.

In larger systems, the question of colour correction becomes important at apertures above $F/3$. Evidently two plates, of different glasses, are needed to achieve it; these plates must be of opposite asphericities and (if they are both to lie in the parallel incoming beams) of strengths inversely proportional to their dispersions. J. G. Baker (1940), in his paper "A family of flat-fielded cameras equivalent in performance to the Schmidt camera", proposed to use two such plates in contact, and A. Warmisham (1941, 1943) has described systems in which each of two spherical mirrors is corrected by a plate placed at its (actual or imaged) centre of curvature.* From the point of view of the general theory of plate-mirror systems, the condition that the plates should be in contact, or that they should each anastigmatize one spherical mirror, appears a somewhat arbitrary limitation, and it is natural to ask whether improved types of two-plate Schmidt-Cassegrain systems may not become available when it is removed. The main object of the present paper is the discussion of this question.

In § 2 we show how, by the use of two spherical mirrors and two aspherical plates, a two-parameter family of flat-fielded anastigmats can be constructed, apochromatic and satisfying the exact sine condition. These systems seem to offer some advantages in ease of construction over the achromatized Baker cameras; they involve only two aspherical surfaces, as compared with three in

* Warmisham does not mention colour correction in the first of these patents, but his examples allow it by a suitable choice of glasses.

the Baker systems (or four in Baker's type D) Further, the aspherized surfaces are both pseudo-flats, which are easier to figure than more sharply curved surfaces.

In §3 we consider the effect of allowing small departures from strict Seidel anastigmatism, such as are required to balance the higher errors over a finite field, and show that the range of practical anastigmats is considerably wider than is suggested by the Seidel theory alone.

§2 ANASTIGMATS

The system to be discussed consists of two aspheric, nearly plane-parallel plates, in general of different glasses, and two spherical or figured mirrors, namely a large concave primary and a small convex secondary. The arrangement is shown in figure 1. The light from the object-surface, supposed infinitely distant, passes through the two plates P , P' , and is reflected from the surface of the concave primary M_1 on to the convex secondary M_2 . There it is again

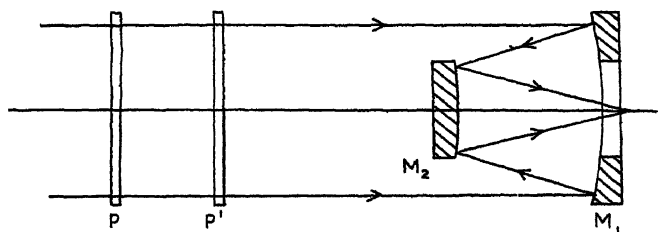


Figure 1.

reflected and comes to a focus near or just behind the primary, which is pierced with a central hole in the latter case to allow the rays to pass. In short, the system is an extension of the Schmidt-Cassegrain type* in which the single aspheric plate of the ordinary Schmidt-Cassegrain is replaced by two plates, at different distances in front of the primary. We shall refer to it as a two-plate Schmidt-Cassegrain system.

The Seidel errors of such a system can be most conveniently investigated by the method of plate-diagram analysis. Let

A , B , Γ , Γ' be the figuring depths on M_1 , M_2 , P , P' respectively, expressed in terms of parabolic correction of M_1 as unit;

σ , σ' the distances of P , P' in front of M_1 , expressed in terms of the paraxial focal length f_1 of M_1 as unit;

ξ the ratio f_2/f_1 of the paraxial focal length (taken positively) of M_2 to that of M_1 ;

$\delta = d/f_1$ the separation between the mirrors, expressed in terms of f_1 as unit;

$q = 1 - \delta$ the "minimum obstruction ratio" for the on-axis pencil.

Then, assuming for the present that the aperture-stop is on the primary, the system possesses the plate-diagram shown in figure 2. From this the aplanatism conditions can be read off in the form

* See *Mon. Not. R. Astr. Soc.* 104, 48-64 (1944) A knowledge of the contents of this paper is assumed in what follows.

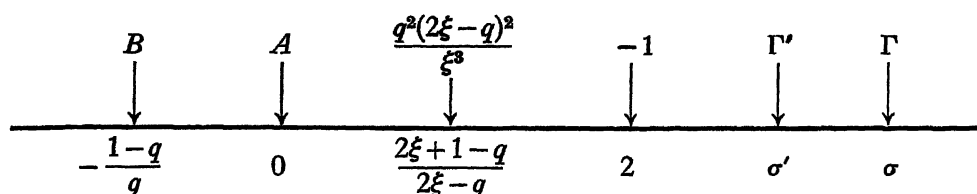


Figure 2.

$$(\text{Spherical aberration}) \quad A + B + \Gamma + \Gamma' + \frac{q^2(2\xi - q)^2}{\xi^3} - 1 = 0. \quad \dots\dots(1)$$

$$(\text{Coma}) \quad -\frac{1-q}{q}B + \sigma\Gamma + \sigma'\Gamma' + \frac{q^2(2\xi - q)(2\xi + 1 - q)}{\xi^3} - 2 = 0, \quad \dots\dots(2)$$

and the astigmatism of the system is measured by the quantity*

$$\chi = \frac{\text{astigmatism}}{\text{thin lens value}} = \frac{-\xi}{4(\xi - q)} \left[\left(\frac{1-q}{q} \right)^2 B + \sigma^2\Gamma + \sigma'^2\Gamma' + \frac{(2\xi + 1 - q)^2 q^2}{\xi^3} - 4 \right]. \quad \dots\dots(3)$$

In these systems, the image is formed at a distance

$$\left(\frac{2\xi - q}{\xi - q} q - 1 \right) f_1 \quad \dots\dots(4)$$

behind the pole of the primary, and the focal length

$$f = \frac{\xi}{\xi - q} f_1. \quad \dots\dots(5)$$

The Petzval curvature of the system

$$\frac{1}{\rho_P} = 2(D - C) = -\frac{1}{f_1} + \frac{1}{f_2} = \frac{1}{f_1} \cdot \frac{1 - \xi}{\xi}, \quad \dots\dots(6)$$

and its distortion coefficient Δ is given by the equation†

$$-8\Delta = \sigma^2\Gamma + \sigma'^2\Gamma' - \left(\frac{1-q}{q} \right)^3 B + \frac{2\xi + 1 - q}{2\xi - q} \frac{(2\xi + 1 - q)^2 q^2 - 4\xi^2}{\xi^3}; \quad \dots\dots(7)$$

when the astigmatism is zero, the value of Δ is not changed by moving the aperture stop.

Two-sphere two-plate systems

Of special practical importance are the systems in which both mirrors are spherical and the colour-error is reduced to a negligible amount. We shall refer to these as the two-sphere two-plate systems. To obtain the aplanatism conditions for such a system, we set $A = B = 0$ in equations (1), (2). We agree to choose the notation so that P denotes the stronger plate and P' the weaker, and we define

$$k = -\Gamma'/\Gamma;$$

* The physical meaning of χ is as follows: at an angle θ off axis in a system of focal length, the astigmatic difference is $\chi\theta^2 f$.

† *Mon. Not. R. Astr. Soc.* 104, 54 (1944), equation (13).

thus $0 < k \leq 1$ and k measures the ratios of the dispersions of the glasses required to give apochromatism. (1), (2), (3) then take the form

$$(1-k)\Gamma = P, \quad \dots\dots (8)$$

$$(\sigma - k\sigma')\Gamma = Q, \quad \dots\dots (9)$$

$$x = \frac{\text{Astigmatism}}{\text{thin lens value}} = \frac{-\xi}{4(\xi - q)} [(\sigma^2 - k\sigma'^2)\Gamma - R], \quad \dots\dots (10)$$

where *

$$\left. \begin{aligned} P &= 1 - \frac{q^2(2\xi - q)^2}{\xi^3}, \\ Q &= 2 - \frac{q^2(2\xi - q)(2\xi + 1 - q)}{\xi^3}, \\ R &= 4 - \frac{q^2(2\xi + 1 - q)^2}{\xi^3}. \end{aligned} \right\} \quad \dots\dots (11)$$

The aplanatism conditions (8), (9) are together equivalent to the equations

$$\Gamma = \frac{P}{1-k}, \quad \sigma' = \frac{1}{k} \left[\sigma - \frac{(1-k)Q}{P} \right], \quad \dots\dots (12)$$

and on substituting from (12) we obtain (10) in the form

$$\begin{aligned} x &= \frac{-\xi}{4(\xi - q)} \left[\frac{\sigma^2 - k\sigma'^2}{1-k} P - R \right] \\ &= \frac{-\xi}{4(\xi - q)P} \left[(Q^2 - PR) - \frac{1}{k} (P\sigma - Q)^2 \right] \\ &= \frac{-\xi}{4(\xi - q)P} \left[\frac{q^2}{\xi^3} (2\xi - 1 - q)^2 - \frac{1}{k} (P\sigma - Q)^2 \right]. \quad \dots\dots (13) \end{aligned}$$

Since this can only vanish if $k > 0$, it follows that the plate-strengths must be of opposite sign in an anastigmat (or near-anastigmat). Equating the astigmatism to zero gives

$$\sigma = \frac{Q}{P} \pm \frac{1}{P} \sqrt{\frac{k}{\xi^3}} q(2\xi - 1 - q). \quad \dots\dots (14)$$

It follows that all the two-sphere two-plate aplanats are given by (12) and all the two-sphere two-plate anastigmats by (12) together with (14). From (12) we have

$$\sigma' - \sigma = \left(\frac{1}{k} - 1 \right) \left(\sigma - \frac{Q}{P} \right) = \pm \left(\frac{1}{\sqrt{k}} - \sqrt{k} \right) \frac{q(2\xi - 1 - q)}{P\xi^{3/2}}. \quad \dots\dots (15)$$

Thus to each choice of the three parameters q, k, ξ , subject to the inequalities

$$0 < q < 1, \quad 0 < k \leq 1, \quad 2\xi - 1 - q > 0, \quad \dots\dots (16)$$

* P, Q and R are seen to be the sign-reversed contributions from the two power-surfaces (excluding figurings) to the spherical aberration ΣA_1 , the coma $\Sigma A_1\sigma$, and the astigmatism $\Sigma A_1\sigma^2$ of the system.

correspond just two anastigmatic systems, namely, those given by

$$\begin{aligned}\sigma &= \frac{Q \pm k^{1/2}q(2\xi - 1 - q)/\xi^{3/2}}{P}, & \Gamma &= \frac{P}{1-k}, \\ \sigma' &= \frac{Q' \pm k^{-1/2}q(2\xi - 1 - q)/\xi^{3/2}}{P}, & \Gamma' &= \frac{P}{1-1/k},\end{aligned}\quad (17)$$

where both the upper signs or both the lower signs are to be taken.*

The two anastigmats (17) have the same plate strengths and the same plate separation, but the order of the plates is reversed on passing from one to the other. The overall length is less when the lower signs are taken; as starting point for an astrographic camera design, therefore, this second system is to be preferred.

Flat-fielded anastigmats

The necessary and sufficient condition that an anastigmat should be flat-fielded in Seidel approximation is that its Petzval curvature should be zero. Thus the flat-fielded anastigmats are the solutions of (12), (14) with $\xi = 1$. In this case the inequalities (16) reduce to

$$0 < q < 1, \quad 0 < k \leq 1,$$

and the second solution (17) to

$$\begin{aligned}\sigma &= \frac{Q - k^{1/2}q(1-q)}{P}, & \Gamma &= \frac{P}{1-k}, \\ \sigma' &= \frac{Q - k^{-1/2}q(1-q)}{P}, & \Gamma' &= \frac{P}{1-1/k},\end{aligned}\quad \dots\dots(18)$$

where now

$$P = 1 - q^2(2 - q)^2, \quad Q = 2 - q^2(2 - q)(3 - q). \quad \dots\dots(19)$$

Thus we have obtained a two-parameter family of flat-fielded anastigmats, apochromatic and with both mirrors spherical, which can in addition be made to satisfy the exact sine condition.†

Distortion in the flat-fielded anastigmats

Setting $B = 0$, $\xi = 1$ in (7), we obtain for the distortion coefficient Δ in the systems (18) the equation

$$-8\Delta = \sigma^3\Gamma + \sigma'^3\Gamma' - \frac{3-q}{2-q}R,$$

where R stands for the quantity $4 - q^2(3 - q)^2$,

$$\begin{aligned}&= \frac{[Q - (\sqrt{k})q(1-q)]^3}{P^2(1-k)} + \frac{[Q - (1/\sqrt{k})q(1-q)]^3}{P^2(1-1/k)} - \frac{3-q}{2-q}R \\ &= \frac{1}{P^2}[Q^3 - 3Qq^2(1-q)^2 + (\sqrt{k} + 1/\sqrt{k})q^3(1-q)^3] - \frac{3-q}{2-q}R;\end{aligned}$$

i.e.

$$\Delta = -\frac{q^3(1-q)^3}{8P^2}(\sqrt{k} + 1/\sqrt{k} - H), \quad \dots\dots(20)$$

* If $2\xi - 1 - q = 0$, i.e. if the mirrors are concentric, the two solutions run together into the single solution

$$\sigma = \sigma' = Q/P, \quad \Gamma = -\Gamma'/k = P/(1-k),$$

a monocentric Schmidt-Cassegrain with achromatized plate

† By essentially the same procedure as that outlined (for the case of a Baker B system) in *Proc. Phys. Soc.* 55, 494 (1943)

where

$$H = \frac{1}{q^3(1-q)^3} \left\{ 3Qq^2(1-q)^2 - Q^3 + \frac{3-q}{2-q} P^2 R \right\} \\ = 3 \frac{Q}{q(1-q)} - \left(\frac{Q}{q(1-q)} \right)^3 - \frac{3-q}{2-q} \left[\left(\frac{Q}{q(1-q)} \right)^2 - 1 \right]. \quad \dots\dots(21)$$

Equation (20) has been derived on the assumption that the aperture-stop of the system is on the surface of the primary mirror M_1 . Since Δ is invariant under stop-shifting in a flat-fielded anastigmat, (20) remains valid when this restriction is removed.

Calculation shows that H is negative for $0.3 \geq q \geq 0.7$. Thus all the anastigmats in this q -range suffer from barrel distortion. Its amount decreases with increasing k ; table 1 shows, for selected values of q , the values Δ_1, Δ_2 of Δ in the two cases $k = \frac{1}{2}, k = 1$; the corresponding values Δ_B in a Baker B system are included for comparison.

Table 1

q	H	Δ_1	Δ_2	Δ_B
0.4	-35.62	-0.187	-0.186	-0.205
0.45	-22.55	-0.177	-0.176	-0.198
0.5	-14.25	-0.167	-0.166	-0.191
0.6	-5.13	-0.145	-0.142	-0.174
0.7	-0.94	-0.120	-0.115	-0.152

We conclude that, in the region $0.4 \leq q \leq 0.7, \frac{1}{2} \leq k \leq 1$, the distortion is nearly independent of k , is slightly less than that of a Baker B system, and is given to within 10% accuracy by the approximate formula

$$\Delta \simeq -0.08q.$$

At 3° off axis this corresponds in the case $q = 0.4$ to a fractional distortion of less than 0.05%.

The Warmisham systems

As a special case we may suppose that $k = q^2(2-q)^2$. Then (18) simplifies to

$$\sigma = 2, \quad \Gamma = 1; \quad \sigma' = \frac{3-q}{2-q}, \quad \Gamma' = -q^2(2-q)^2, \quad \dots\dots(22)$$

and we obtain a one-parameter family of flat-fielded anastigmats in which the outer plate is simply the Schmidt plate of M_1 in its normal position, and the inner is the image in M_1 of the Schmidt plate of M_2 .^{*} Each plate thus anastigmatizes one mirror. The restriction $k = q^2(2-q)^2$, while unnecessary from our present point of view, is not a very serious limitation in practice. For example, to obtain $k = \frac{1}{2}$ and so allow on-axis colour correction with glasses chosen from an ordinary crown-flint pair, we need to choose $q = 1 - \sqrt{(1 - \sqrt{k})} = 0.459$. This brings the image to a very convenient position $0.31f_1$ behind the front

^{*} The Gauss image in M_1 of the centre of curvature of M_2 is at a distance

surface of the primary. The focal length f is then $1.85f_1$ and the plate separation $\sigma - \sigma'$ is $0.35f_1 = \frac{1}{5}f$.*

When the chosen value of q is not close to 0.46, the general two-sphere anastigmats provide alternatives which offer some advantages over the corresponding Warmisham systems. When q is larger than 0.46 they offer the possibility of reducing the plate-strengths by giving to k a smaller value than $k_w(q) = q^2(2-q)^2$, while preserving full achromatism on axis. This reduction in plate-strength is obtained at the cost of a small increase in the overall length of the system. When q is smaller than 0.4, on the other hand, they allow on-axis achromatism to be maintained with conveniently available glass-types by giving to k a value greater than $k_w(q)$. This involves an increase in the plate strengths. A survey of the situation is provided by table 2, in which the constants σ , σ' , Γ , Γ' of anastigmats with $k = \frac{1}{2}$ are compared, for selected values of q , with those σ_w , σ'_w , Γ_w , Γ'_w of the corresponding Warmisham systems.

Table 2

q	σ	σ'	Γ	Γ'	σ_w	σ'_w	Γ	Γ'_w	$k_w(q)$	$ f/f_1 = 1(1-q)$
0.3	1.944	1.743	1.480	-0.740	2	1.588	1	-0.260	.2601	1.4286
0.4	1.973	1.685	1.181	-0.590	2	1.625	1	-0.410	.4096	1.6667
0.5	2.025	1.620	0.875	-0.438	2	1.667	1	-0.563	.5625	2.0000
0.6	2.108	1.532	0.589	-0.294	2	1.714	1	-0.706	.7056	2.5000
0.7	2.248	1.384	0.344	-0.172	2	1.769	1	-0.828	.8281	3.3333

§3 THE TWO-SPHERE TWO-PLATE APLANATS

The conditions (1) spherical aberration = 0, (2) coma = 0, (3) astigmatism = 0, (4) Petzval curvature = 0 of the Seidel approximation are to be replaced in practice by the conditions that these quantities should be small. This relaxation has only a small effect on the range of available systems so far as conditions (1), (2) and (4) are concerned, and we can therefore obtain a satisfactory survey of the useful systems by keeping the spherical aberration, coma and Petzval curvature strictly zero. With condition (3) the case is different; the admission of a small amount of astigmatism widens the range of possible systems considerably. It follows that, in making a survey of the useful two-sphere two-plate systems, it is necessary to consider not merely the Seidel anastigmats but the larger class of Seidel aplanats with zero Petzval curvature.† When this is done, an interesting result appears. Fixing q and k , let us consider the changes in the system as σ varies. To each value of σ corresponds a unique aplanat; for two values of σ the astigmatism vanishes and we recover the two anastigmats of §2. In between these two values of σ the astigmatism is negative and small. Thus we obtain not two, but a whole range of useful systems. At the σ -point where the negative astigmatism reaches its worst value, the plates cross over, and when σ is greater than this value the weaker σ' -plate is the farther from the primary. At the cross-over value of σ , the system is equivalent to a two-sphere one-plate aplanat with achromatized plate.

* Of course it would be simpler, if our object were merely to derive the Warmisham system to begin by combining a positive and a negative Schmidt system of equal mirror-curvatures. On imaging the plate of the negative system through into object-space we at once obtain (22).

† Not with flat "best field" in the Seidel sense, for reasons similar to those obtaining in the case of a two-sphere one-plate aplanat (*Mon. Not. R. Astr. Soc.* **104**, 61-62 (1944)).

To prove these statements, we set $\xi=1$ in (12) and obtain the equations of the systems to be discussed in the form

$$\Gamma = \frac{P}{1-k}, \quad \dots\dots(23)$$

$$\sigma' = \frac{1}{k} \left[\sigma - \frac{(1-k)Q}{P} \right], \quad \dots\dots(24)$$

where P, Q are given by (19) Equation (13) now becomes

$$\chi = -\frac{1}{4(1-q)P} \left[q^2(1-q)^2 - \frac{1}{k}(P\sigma - Q)^2 \right]. \quad \dots\dots(25)$$

From (24) we have, as before,

$$\sigma' - \sigma = \left(\frac{1}{k} - 1 \right) \left(\sigma - \frac{Q}{P} \right). \quad \dots\dots(26)$$

For given q, k the values of Γ and $\Gamma' = -k\Gamma$ are fixed by (23), while σ' is given in terms of q, k and σ by (24). Thus the system is completely determined by choice of the parameter σ , which measures the distance of the stronger plate from the primary. Variation of σ is an operation which is very easily carried out experimentally; it consists merely in moving the plates P and P' along the axis of the system, the new position of P' being determined by (24). In this way the Seidel astigmatism of the system can be varied more or less at will. Of course the higher aberrations are also disturbed to some extent.

If, starting from the anastigmat (18), we increase σ by moving the plate away from the primary, it follows that from (26) the plate P' moves out also and decreases its distance from P , overtaking and passing it when $\sigma = Q/P$. From (25) we see that the ratio $\chi = \text{astigmatism}/(\text{thin lens value})$, which is negative when σ lies between the two anastigmat values $[Q \pm \sqrt{(k)q(1-q)}]/P$, attains its greatest negative value $-q^2(1-q)/4P$ when $\sigma = Q/P$, that is, when the plates are in the act of crossing over. Now for this value of σ the system is simply a two-sphere one-plate aplanat with achromatized plate. The astigmatism of such a system is so small when $q \leq 0.5$ that it is for all practical purposes anastigmatic, as will be seen from table 3.

Table 3

q	$\chi = -q^2(1-q)/4P$
0.3	-0.0213
0.35	-0.0299
0.4	-0.0407
0.45	-0.0542
0.5	-0.0714

It follows that *throughout the σ -range between the two anastigmat values of σ , and for a short distance outside this range at either end, the two-sphere two-plate aplanats possess less astigmatism than a Schmidt-Cassegrain aplanat with the same value of q , and hence are for all practical purposes anastigmatic.* We have supposed $q \leq 0.5$; this covers the astrographically valuable types.

Distortion

The value of the distortion coefficient now depends on the position of the aperture-stop. Let Δ_z denote its value when the stop is at a distance ζf_1 in front of the primary. When $\zeta=0$, Δ_z reduces to the coefficient Δ of equation (7). In the general case we have *

$$\begin{aligned}\Delta_z &= \Delta + \zeta f_1 \left(\frac{3}{8} \sum_P \alpha x^2 - \frac{1}{2} \sum_A \kappa \right), \\ &= \Delta + \zeta \left(\frac{3}{8} \sum_P A \sigma^2 - \frac{1}{2} f_1 \sum_A \kappa \right).\end{aligned}\quad \dots\dots(27)$$

Here $\sum_A \kappa=0$, since the Petzval curvature is zero, while

$$\sum_P A \sigma^2 = -4\chi f_1/f,$$

by (3) and (5). Thus

$$\Delta_z = \Delta - \frac{2}{3}\chi\zeta.\quad \dots\dots(28)$$

In practice, the value of ζ will usually be near to 1. For, assuming the system to be intended for ordinary astrographic use, it is very desirable to obtain as even an illumination of the field as possible, with the least possible sacrifice

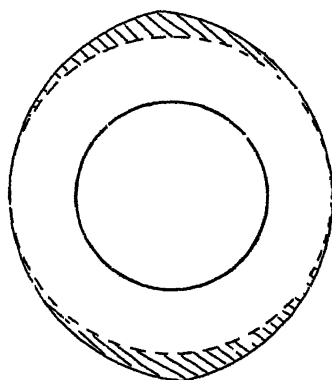


Figure 3.

of aperture. This can be done by constructing the two plates and the large mirror all of the same diameter and inserting an aperture-stop, of slightly smaller diameter, half-way between the outer plate and the primary.† The diameter of this stop is chosen so that, at the edge of the designed field of 6° or 8° diameter, the outer edge of the effective aperture is not affected by vignetting.‡

In systems intended for astronomical search-photography, for example the detection of novae, it is better not to sacrifice light-gathering power by introducing an aperture-stop, but to allow the outer plate and the primary mirror to vignette

* *Mon. Not. R. Astr. Soc.* 104, 62 (1944), equation (28).

† It is tempting to pick out a class of "distortion-free" systems by the criterion that $\Delta_z=0$ when $\zeta=\frac{1}{3}\sigma$. But such a classification would have little practical value, because the empirical changes in the Petzval sum which are needed, at aperture near $F/2$, to flatten the field to the best advantage cause changes in Δ_z too great to be corrected by a small shift of the aperture-stop.

‡ Since $\cos^2 3^\circ = 1 - 1/400$ nearly, the illumination over a 6° diameter field will then fall by $\frac{1}{4}\%$ from the centre to the edge of the field.

against each other. This applies with the greater force because the parts of the aperture which have to be sacrificed when uniform lighting is insisted upon are valuable regions of small error slope", which contribute towards the central nucleus of the images near the edge of the field. These regions are shaded in figure 3, which shows the general appearance of the effective aperture.*

Two main types of system can now be selected as especially promising for astronomical application:

Type A. Systems in which σ is a little below the cross-over value Q/P . In these systems, the plates are near together but not in contact. Their separation makes it possible to satisfy the exact sine condition, and so eliminate the most important error (higher coma) of a two-sphere one-plate aplanat with achromatized plate, at a negligible cost in off-axis colour due to the shearing of the plates. Table 4 gives the Seidel data of six examples of this type; the fourth and fifth are illustrated in figure 4.

Table 4

q	k	σ	σ'	Γ	Γ'	x_{max}	Δ	f/f_1
0.3	0.5	1.9	1.6553	1.4798	-0.7399	+0.0104	+0.1331	1.4286
0.35	0.5	2.0	1.8029	1.3330	-0.6665	-0.0100	-0.0046	1.5385
0.4	0.5	2.0	1.7398	1.1808	-0.5904	-0.0074	-0.0922	1.6667
0.4	0.6	2.1	1.9932	1.4760	-0.8856	-0.0301	+0.0430	1.6667
0.45	0.5	2.0	1.6638	1.0270	-0.5135	-0.0015	-0.1647	1.8182
0.45	0.6	2.1	1.9425	1.2837	-0.7702	-0.0325	-0.0720	1.8182

Type B. Systems of short overall length. To obtain these we have to choose σ as small as possible. The practical lower limit to σ is determined by

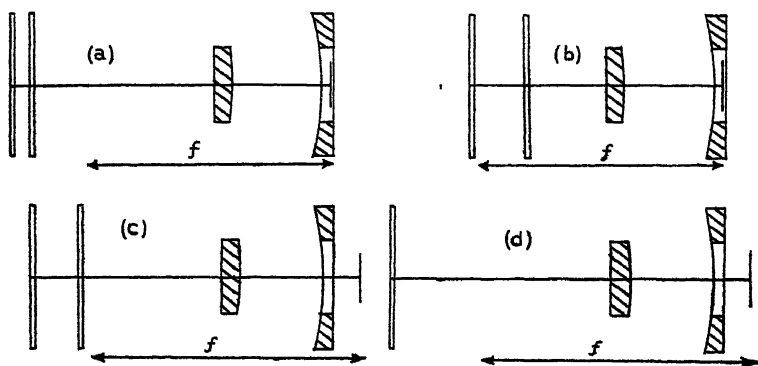


Figure 4.

- (a) Type A, $q=0.4$, $k=0.6$, $\sigma=2.1$.
 (b) Type B, $q=0.4$, $k=0.6$, $\sigma=1.67$.
 (c) Type A, $q=0.45$, $k=0.5$, $\sigma=2.0$.

- (d) Baker B system, $q=0.45$, for comparison; in this system the primary mirror is aspherical.

the astigmatism, which increases fairly rapidly when σ is decreased below its smaller anastigmat value. For a thin lens of twelve inches aperture the diameter

* The case actually illustrated is that of a system, of nominal aperture-ratio $f/2.5$, in which $\sigma=2$, $q=0.4$ and the diameter of the secondary mirror is half that of the primary, at 3° off axis. The effective aperture-ratio is $f/3$.

of the astigmatic confusion-circle 3" off axis is 0.015 inch at best focus. If we adopt one tenth of this value as the upper limit to the admissible Seidel astigmatism, we are led to impose the condition

$$|\chi| \leq 0.1,$$

which then determines the least permissible value of σ . Table 5 gives the Seidel data of five examples of type B systems; the third example is illustrated in figure 4.

Table 5

q	h		σ'	Γ	Γ''	χ	A	f/f_1
0.3	0.5	1.6657	1.1866	1.4798	-0.7399	0.1000	0.2821	1.4286
0.4	0.5	1.7255	1.1908	1.1808	-0.5904	0.1000	0.0663	1.6667
0.4	0.6	1.6745	1.2840	1.4760	-0.8856	0.1000	0.2431	1.6667
0.45	0.5	1.7614	1.1866	1.0270	-0.5135	0.1000	-0.0276	1.8182
0.45	0.6	1.7065	1.2868	1.2837	-0.7702	0.1000	0.1160	1.8182

REFERENCES

- BAKER, G. J., 1940. *Proc. Amer. Phil. Soc.* 82.
 LINFOOT, E. H., 1944. *Mon. Not. R. Astr. Soc.* 104
 WARMISHAM, A., 1941. B.P. 541,650, 1943. B.P. 551,082.

ON GRÜNEISEN'S EQUATION FOR THERMAL EXPANSION

By WILLIAM HUME-ROTHERY, F.R.S.,
Oxford

MS. received 15 January 1945

ABSTRACT. The practical use of Grüneisen's relations as a means of calculating changes in volume from changes in thermal energy is discussed. The relation is often misquoted in an inaccurate and approximate form, and the correct form is described. The methods used for deducing the constants involved are discussed critically, and a simple and accurate method is described. For the cubic metals silver, copper and aluminium there is a very good agreement between the observed and calculated changes in volume between the absolute zero and a temperature of the order 2/3 of the melting point on the absolute scale, whilst for iron there is good agreement up to the temperature at which the magnetic transformation begins. For the hexagonal metals magnesium and zinc, the equation fails at very low temperatures if the substance has a negative coefficient of expansion in one direction (e.g. zinc), but is otherwise fairly satisfactory for the calculation of changes in volume up to a temperature of the order 2/3 of the absolute melting point.

1. INTRODUCTORY

IN some recent work by the author, it was necessary to calculate the lattice spacings of metals in ranges of temperature for which the coefficient of expansion was unknown. This led naturally to the use of Grüneisen's equation connecting changes in volume with changes in thermal energy. Examination showed that confusion existed both as to the exact nature of

Grüneisen's relation and as to the accuracy which might be expected from its use. In this paper we examine these two points, and we present calculations for some metals for which both the coefficients of expansion and the specific heats are known accurately. We are concerned throughout with the practical use of Grüneisen's relation as a means for calculating changes in volume, and we do not attempt to discuss the theoretical implications.

§ 2. GRÜNEISEN'S RELATION

Grüneisen's relation is commonly described by saying that when the temperature of a body is altered, the change in volume is proportional to the change in thermal energy, provided that the substance is one for which the specific heat can be expressed in terms of a single Debye characteristic temperature, Θ . This definition lacks precision, and has been responsible for much confusion. It has, for example, led some authors to conclude that Grüneisen's relation implies a constant value for the ratio C_p/α , where α is the coefficient of expansion and C_p is the atomic heat at constant pressure; as we show later, this is a quite unjustifiable approximation of what Grüneisen's work really involves. Of the more precise definitions of Grüneisen's relation, we shall adopt the modification used by Simon and Vohsen. This may be expressed in the form

$$\frac{V_T - V_0}{V_0} = \frac{E_T}{Q_0 - kE_T}, \quad \dots\dots(1)$$

where V_0 and V_T are the volumes at the absolute zero and at $T^\circ \text{K}$. respectively, and $E_T = \int_0^T C_v dT$. Q_0 is a constant whose derivation is discussed below, and k is a constant which is equal to $(\gamma + 2/3)$, where

$$\gamma = -V \frac{\partial V}{\partial T} / C_v \frac{\partial V}{\partial p}.$$

An alternative form of the equation replaces the constant k by an expression involving the coefficients m and n of a hypothetical law of force in which the attraction and repulsion are proportional to terms of the type $1/r^m$ and $1/r^n$ respectively.

The magnitudes involved in equation (1) are such that for most metals Q_0 is of the order 10^4 to 10^5 cal., whilst k is a small number of the order 1 to 3. E_T is of the order 1000 at room temperature, and increases by roughly 600 for each 100° rise of temperature. At low temperatures, the results depend essentially on E_T and Q_0 , but at high temperatures the terms kE_T becomes increasingly important.

The question of the validity of equation (1) has involved a great deal of confusion, and some of the main causes of difficulty may be summarized as follows :

(a) The term Q_0 may be determined by several methods which are referred to in § 5 below. Some of these are at the best only approximations, whilst others involve physical constants which may only be known approximately. This means that errors of the order 5 to 20% may be introduced at the outset, through an incorrect choice of Q_0 .

(b) The fact that the theoretical basis of Grüneisen's relation involves the assumption of a single Debye characteristic temperature has led to a general tendency to calculate E_T by means of Debye functions. This has led to confusion in two ways. In the first place, specific heats do not always follow Debye functions accurately, and in some cases there is a tendency to choose the value of Θ so that it gives the correct specific heat at the low temperatures, where the variation in C_v is most interesting. For the calculation of changes of volume, however, it is much more important for the specific heat to be correct at high temperatures, because at low temperatures both C_v and α are small, and relatively large errors may occur without much effect on the change in volume. In the second place, the uncertainty in the value of Q_0 is not always realized. A procedure sometimes adopted is to deduce Q_0 by one of the approximate relations given below, and to obtain a value which is in error by a few per cent. When the value of Q_0 is combined with the correct* value of Θ , the calculated changes in volume are naturally wrong by an amount corresponding with the error in Q_0 . Instead of this being recognized, a fictitious characteristic temperature Θ' is then introduced, and is used to calculate a fictitious value of E_T which gives a better agreement with the volume changes. In fact it is hardly unfair to say that an incorrect value of Q_0 is made an excuse for introducing a fictitious Θ' , and then the combination of the two is claimed as a proof of the correctness of Grüneisen's relation. It is clear that, logically, this is very unsatisfactory, whilst for the purposes of the practical use of Grüneisen's relation it is of course valueless, since the fictitious Θ' can be deduced only when the changes in volume are known. It is this which makes the otherwise interesting work of Nix and MacNair (1941) useless for the present purpose. These authors show convincingly that for some metals the coefficients of expansion agree with equation (1), provided that Q_0 is chosen arbitrarily, and that E_T is calculated from the Debye function with a suitably chosen value of Θ . It was, however, sometimes necessary to make slight adjustments to the values of Θ obtained from specific-heat data, and consequently, although the conclusions were of considerable interest, they did not constitute a proof of the relation between changes of volume, and the real value of $E_T = \int_0^T C_v dT$.

(c) Apart from the above points, there has been a tendency to confuse the issue by considering not the actual changes in volume, but the calculated and observed mean coefficients of expansion over relatively wide ranges of temperature. It may, for example, at first appear satisfactory to state that the calculated and observed mean coefficients of linear expansion are 2.0×10^{-5} and 2.2×10^{-5} respectively, but the difference is equivalent to a 10% error in changes of length and a 30% error in changes of volume.

If we ignore the theoretical implications of Grüneisen's relation, and consider it solely from the practical point of view, it will be seen that there are two questions to be decided. It is necessary to examine first of all whether changes in volume are related to E_T by a relation of the type of equation (1), provided that Q_0 and k are chosen suitably, and secondly, to see how Q_0 and k are to be

* We are here assuming that the substance is one to which a Debye function applies accurately.

determined, the correct choice of Q_0 being the real essential. In §§ 4 and 5 we shall examine these two points.

§ 3. ACCURACY OF EXPERIMENTAL DATA

For critical examination of equation (1) it is necessary to consider the accuracy of the experimental work on thermal expansion, and it is convenient to consider first the work on isotropic (cubic) structures. Measurements of the coefficients of expansion of massive metal at high temperatures are often unreliable owing to the effects of gas or other impurities in the metal. Thus, variations in the thermal expansion of copper have been ascribed to dissolved hydrogen acting on traces of oxide with the liberation of steam, which, on blowing out of the metal, may produce an abnormal expansion. These difficulties are overcome by the use of high-temperature x-ray technique, which enables the true lattice spacings to be measured, irrespective of the presence of blowholes. Lattice-spacing measurements at room temperatures should be consistent to 1 part in 50,000, but the high-temperature work involves many possible errors (e.g. calibration of thermo-couples, differences in temperature during exposure, etc.), and the accuracy is undoubtedly lower. Thus Wilson (1942), in repeating very careful work on the lattice spacings of aluminium, found differences of the order 1 part in 14,000. It seems probable, therefore, that even in good quality work on simple substances, giving satisfactory high angle x-ray reflections, the lattice-spacing measurements may sometimes contain errors increasing from 1 part in 20,000 at moderate temperatures to 1 part in 10,000 or 1 part in 5000 at high temperatures.

With non-cubic substances, the coefficient of expansion generally varies with the direction relative to the crystal axes. Measurements on massive polycrystalline rods are often valueless, because of preferred orientation of the crystals.* This difficulty can be overcome by measuring the expansion of a cube in three directions, but such measurements are uncommon. Measurements on massive single crystals should be satisfactory, but some of the early work in this connexion is very inaccurate, presumably owing to accidental strain of the crystals. It is also necessary to distinguish carefully between investigations designed expressly to measure coefficients of expansion and those in which data on expansion were obtained as a by-product of other work (e.g. of compressibility). In the latter type of work the experimental arrangements may be unsuitable for accurate measurements of expansion. The x-ray method is again satisfactory, but since it involves the determination of two parameters, the error is at least twice as great as for cubic substances, unless the structure is such that a high-angle line of the form $(h, 0, 0)$ enables one parameter to be obtained independently of the choice of axial ratio.

For temperatures below 0° C. there does not yet seem to be an x-ray camera designed for detailed work over a range of temperatures, but fortunately there is accurate work on polycrystalline specimens of the cubic metals referred to below, and on single crystals of the non-cubic metals.

We shall not discuss in detail the accuracy of specific-heat data. For some

* Bars of drawn and annealed metal nearly always show marked preferred orientation, whilst cast bars may show some preferred orientation.

of the metals dealt with below, the specific heats at room temperatures are probably reliable to 1 part in 300, but for many metals the data are uncertain by at least 1%. At low temperatures the accuracy is less, but as the specific heats are smaller, the effect is not serious for the calculation of changes in volume. At high temperatures there is a serious falling off in the accuracy with which specific heats are known, and it is seldom that results are reliable above 500° or 600° c.

§ 4. THE ACCURACY OF GRÜNEISEN'S EQUATION (1)

In order to examine the accuracy of equation (1) it is necessary to calculate E_T . For the present purpose it seemed better to discard Debye functions entirely, since their use serves only to introduce the additional complication of the correct choice of Θ . The policy adopted has been to use the actual experimental values of C_p , and to determine the value of $\int_0^T C_v dT$ by graphical integration. The accuracy of the integration is fully as great as is justified by the errors in C_p . The value of $(C_p - C_v)$ can be obtained from the relation (see W. Nernst and F. A. Lindemann, 1911)

$$C_p - C_v = C_p^2 T A,$$

where T is the absolute temperature, and the constant A is given by the relation

$$A = \frac{(3\alpha)^2 V}{\chi C_p^2},$$

α being the linear coefficient of expansion, V the atomic volume, and χ the compressibility. To a first approximation A can be assumed to be independent of temperature, and since at the lower temperatures $(C_p - C_v)$ is small, errors in A are important only at the higher temperatures. The assumption of a constant value of A at the higher temperatures may legitimately be criticized, but on the one hand it does not seem justifiable to examine this point further until the specific heats are known to a higher degree of accuracy, and, on the other, the method used has the advantage of simplicity, and gives the values of E_T in a straightforward way from the C_p data.

The values of k in equation (1) have been taken from the *Handbuch d. Physik*, vol. 10, and as the term kE_T is relatively small at low temperatures, errors in the assumed values of k affect the results only at the higher temperatures.

It is convenient to separate the results for cubic and non-cubic metals. Examination suggested that the data were known accurately for the metals detailed below. References to the original sources are given in the Appendix, where the accuracy of some of the results is also discussed.

Cubic metals. Tables 1, 2, 3, 4 show the results obtained for silver, copper aluminium and iron. In these, and later tables, the quantity calculated is the volume, Ω , per atom in (crystal angstroms)³. Ω is thus equal to one-quarter the volume of the unit cell of the face-centred cubic structure, and one-half the volume of the unit cell of the body-centred cube. In all cases the observed value of the lattice spacing at room temperature is taken as the starting point, and from this the value of Ω at room temperature is deduced, and is used to calculate Ω_0 in equation (1). The values of Ω at different temperatures are then calculated, and are compared with the observed values. At low temperatures,

these "observed values" are based on the coefficients of expansion of massive metal, and at high temperatures the results of accurate measurements of lattice spacings are used. In the case of iron, however, there is no accurate high-temperature x-ray work, and the "observed" values of Ω have been calculated from the coefficients of expansion, which are fortunately known accurately for this metal.

The figures in table 1 show that for silver the observed and calculated values of Ω are in excellent agreement over the whole range 20 to 973° K. Between 20 and 90° K. the calculated increase in Ω is 0.031 Å³ as compared with the

Table 1 Silver

Temperature (° K.)	Volume per atom (Å ³)	
	calculated	observed
973	17.725	17.718
773	17.469	17.468
572	17.237	17.241
273	16.929 value assumed	16.929
171	16.837	16.838
90	16.773	16.773
20	16.742	16.739
$Q=108,010. \quad k=3.07.$		

Table 2. Copper

Temperature (° K.)	Volume per atom (Å ³)	
	calculated	observed
1044	12.260	12.257
944	12.179	12.183
773	12.053	12.054
573	11.903	11.909
273	11.726 assumed	11.726
172.5	11.670	11.672
87.5	11.632	11.635
20	11.621	11.623
$Q=117,000. \quad k=2.63.$		

observed value of 0.034 Å³. The difference is equivalent to an error of roughly 10% in $\Delta\Omega$, and hence of about 3% in the coefficient of *linear* expansion, and this is probably within the experimental accuracy at these low temperatures, where the changes in volume are very small. The results of table 1 suggest that Grüneisen's equation (1) is quantitatively exact over a range of 1000° K. to within the limits of accuracy now available.

Table 2 shows the data for copper, but for the present purpose this metal is not so satisfactory as silver, because the thermal expansion of massive metal in the range 0 to 500° C. does not agree exactly with the expansion of the lattice as measured by x-ray methods. As will be seen from table 2, Grüneisen's equation

is in fairly satisfactory agreement with the data over the range 20 to 1044° K. There are slight differences on opposite sides of 273° K. which cannot be allowed for by changing the value of Q_0 , and these are the result of the above-mentioned discrepancy between the results of the two methods of measuring the thermal expansion. If the expansion above 0° C had been calculated from the data for massive bars of copper (see Appendix) a slightly different value of Q_0 would

Table 3. Aluminium

Temperature (° K.)	Volume per atom (Å ³)	
	calculated	observed
773	17.149	17.163
673	16.997	17.000
573	16.854	16.852
473	16.719	16.714
373	16.592	16.590
273	16.4744 assumed	16.4744
193	16.391	16.391
83	16.306	16.305
20	16.293	16.290

$$Q=84,110. \quad k=2.84.$$

Table 4. Iron

Temperature (° K.)	Volume per atom (Å ³)	
	calculated	observed
958	12.098	12.066
876	12.033	12.019
774	11.963	11.958
678	11.903	11.902
577.5	11.845	11.845
478	11.793	11.792
376	11.745	11.744
293	11.709 assumed	11.709
225	11.682(3)	11.682(5)
131	11.652	11.653
91.5	11.643	11.643
73	11.640	11.640
20	11.638	11.638

$$Q=171,000. \quad k=2.27.$$

have given a better agreement over the range 0 to 553° K., at the expense of a worse agreement at higher temperatures. It does not, however, seem justifiable to discuss the figure in further detail until the causes of the differences between lattice expansion and bulk expansion are understood.

The figures in table 3 show that, for aluminium, equation (1) is confirmed fairly satisfactorily from 20 to 673° K., but at 773° K. the observed expansion is greater than the calculated value by an amount which should exceed the

experimental error, and it seems probable that genuine deviations from equation (1) set in at from two-thirds to three-quarters of the temperature of the melting point on the absolute scale. At 473°K. there is a difference of 0.005 \AA^3 between the calculated and observed values of Ω . This is equivalent to 1 part in 10,000 in the measurement of the lattice spacing, and this is probably somewhat greater than the experimental error, since the work of Wilson (1942) (see Appendix) was carried out in great detail in this range. The specific heats are, however, only known to two significant figures over most of the temperature range, and it can hardly be said that the above difference is outside the experimental error. As in the case of silver, the assumed value of Q_0 gives slightly too small a change in volume between the two lowest temperatures. A small difference in the assumed value of Q_0 would give a better agreement in the range 0 to 273°K. , at the expense of a slightly worse agreement at higher temperatures, but it does not seem profitable to discuss this until the coefficients of expansion at low temperatures, and the specific heats at all temperatures, are known more accurately.

Table 4 shows the results for iron, and this metal is of particular interest, because its specific heat cannot be explained in terms of a single Debye characteristic temperature. It will be seen that, in spite of this, there is an extremely good agreement* between the calculated and observed values of Ω in the range 20 to 774°K. This suggests that equation (1) does not depend on the assumption of a single Debye characteristic temperature, and that the failure of equation (1) for other metals at high temperatures is the result of an effect other than a divergence of the specific heat from a simple Debye function. In the case of iron it will be seen that at temperatures above 774°K. the calculated values of Ω are larger than those observed, and although the difference is not very great† at 958°K. , it suggests that the destruction of the ferro-magnetism involves energy changes which are not related to changes in volume by means of equation (1).

Non-cubic metals. Table 5 shows the calculated and observed values of Ω for magnesium, which crystallizes in the close-packed hexagonal structure with an axial ratio ($c/a = 1.62354$) not very different from that for close-packed spheres ($c/a = 1.6330$). The agreement between the calculated and observed values is clearly not so good as for the cubic metals. From 20 to 473°K. the differences are within the experimental error,‡ and the difference only becomes really pronounced above 773°K. , where the observed expansion is clearly greater; the behaviour here resembles that for aluminium. In the range of temperature below 673°K. there is a suggestion of a systematic departure from equation (1) which could not be corrected by adjusting the value of Q_0 , but the differences are of the same order as the experimental accuracy, and more reliable values for the coefficients of expansion of single crystals at low temperatures and of the specific heats are clearly desirable before discussing the matter further.

* It is to be emphasized that the interest here is that equation (1) holds not merely where a fictitious characteristic temperature is used to calculate a fictitious E_T , but also when the real value of E_T is inserted.

† Reference should be made to the Appendix for difficulties in estimating the specific heats of iron.

‡ The coefficient of expansion of magnesium is so high that in the range 273 to 373°K. an error of 1° changes Ω by approximately 0.002 . An error of 0.01 in Ω is equivalent to one part in 2300 in the volume, or one part in 7000 in the lattice spacings. Since two parameters have to be determined, this is hardly outside the experimental error.

Table 6 shows the results for zinc, which is of interest as a highly anisotropic metal,* which at very low temperatures shows a negative coefficient of expansion perpendicular to the hexagonal axis. In this case equation (1) fails at the lowest temperatures. There is a fair agreement with equation (1) over the range 133 to 573° K., but a critical test is, unfortunately, not possible. The data for high temperatures are from the lattice-spacing measurements of Owen and

Table 5. Magnesium

Temperature (° K.)	Volume per atom (Å ³)	
	calculated	observed
873	24·349	24·379
773	24·099	24·094
673	23·864	23·847
573	23·643	23·630
473	23·434	23·437
373	23·238	23·246
273	23·0561 assumed	23·0561
193	22·923	22·924
90	22·787	22·790
20	22·754	22·753

$$Q=77,660. \quad k=2.18.$$

Table 6. Zinc

Temperature (° K.)	Volume per atom (Å ³)	
	calculated	observed
623	15·602	{ 15·580 15·586
573	15·519	{ 15·502 15·508
473	15·365	{ 15·358 15·364
373	15·221	{ 15·217 15·223
273	15·0861 assumed	15·0861
173	14·959	14·954
133	14·914	14·905
93	14·875	14·860
20	14·833	14·813

$$Q_0=70,750. \quad k=2.68.$$

Yates (1934) (see Appendix), whose value for Ω_{273} is 15·080 as compared with what seems to be the more accurate value of 15·0861 of Jette and Foote (1935). If the difference between the two values is due to a constant factor such as the calibration of a camera, then the values of Owen and Yates require correction by 0·006 to make them comparable with the value assumed (15·0861) for calculation in table 6. If, on the other hand, the difference is due to an error in one

* The structure is close-packed hexagonal with axial ratio $c/a=1.8563$ (Jette and Foote, 1935) at 25° C.

individual experiment of Owen and Yates, the remaining figures of these investigators should be left unaltered. It is for this reason that two values are given at each temperature above 273° K. in table 6, and it is clearly unjustifiable to discuss the results in greater detail until the data are more certain.

From tables 1 to 6 we may conclude that, except for the magnetic transformation in iron, the data for the cubic metals are in very good agreement with equation (1) up to a temperature of the order $\frac{2}{3}$ to $\frac{3}{4}$ of the melting point on the absolute scale. For the non-cubic metals, the equation fails at very low temperatures if the substance has a negative coefficient of expansion in one direction, but is otherwise fairly satisfactory until a temperature of the order $\frac{2}{3}$ of the absolute melting point is reached, although the accuracy of the data is not yet sufficient for a critical test.

§ 5. THE CHOICE OF THE CONSTANT Q_0

From the preceding sections it will be appreciated that the correct choice of Q_0 is all-important if Grüneisen's equation is to give accurate results. The usual relation given in this connexion is

$$Q_0 = \frac{V_0}{\gamma\chi_0}, \quad \dots\dots(2)$$

where γ is the constant referred to in § 1, and equals $-V\frac{\partial V}{\partial T}/C_p\frac{\partial V}{\partial p}$. This method is extremely unsatisfactory because it makes Q_0 inversely proportional to γ , and the values of γ are often very uncertain. Apart from this, the value of χ_0 , the compressibility at the absolute zero, is usually uncertain by several per cent, and it is not surprising that the values of Q_0 obtained in this way require a fictitious characteristic temperature before they are in agreement with the changes in volume.

An alternative method, given by Simon and Vohsen (1928), is to write

$$Q_0 = \frac{C_p}{3\alpha} + 2kE_T, \quad \dots\dots(3)$$

where C_p and α are the values of the specific heat and coefficient of expansion at room temperature. This relation is obtained, as an approximation only, by writing equation (1) in differential form. In principle, the method is much more satisfactory than that involving equation (2), since the term $2kE_T$ (where $k = \gamma + \frac{2}{3}$) is now only a relatively small correction to the value of $C_p/3\alpha$, and the uncertainty in the value of Q_0 is much less serious. Unfortunately, although justified for the purpose of Simon and Vohsen, the approximation is unjustified for high accuracy at room temperature, and it can readily be shown that the values of Q_0 obtained from equation (3) depend on the purely arbitrary value of the temperature (20° C.) at which C_p , α and E_T are taken, even though the substance is one which obeys equation (1).

It will be seen that both of the relations (2) and (3)* require a knowledge of α at some temperature. Consideration will show that if a value of α is known, we can determine Q_0 directly from equation (1) without any approximation. This can be done either by writing equation (1) in differential form, and solving

* γ involves α through the term $\partial V/\partial T$.

the equation for Q_0 , or alternatively by simple methods of trial and error. We have found the latter procedure more convenient, as it provides the constants required for later calculations, and it is convenient for computation. If, for example, the value of α is known at 273°K ., we may assume a reasonable value of Q_0 , and use this to calculate Ω_{253} and Ω_{293} and then adjust Q_0 until the expansion between 253°K . and 293°K . agrees with the known coefficient of expansion at 273°K .

This method is simple and straightforward, and in cases where the expansion obeys equation (1), the method gives correct values of Q_0 , subject only to the errors in the assumed values of α . This uncertainty must not be minimized, because there are many substances for which the coefficients of expansion are uncertain by a few per cent. The direct method of obtaining Q_0 is, however, much more satisfactory than those based on equations (2) or (3), because these involve not only the uncertainty in α , but also other approximations or quantities which are liable to considerable error.

If the substance dealt with is one to which equation (1) applies only approximately over a restricted range of temperatures, then the value of Q_0 obtained by the direct method will be the best possible value for use in the ranges of temperature adjacent to that at which α is known, and will thus be the best value for extending the value of α to new temperatures. We suggest, therefore, that this direct method of obtaining Q_0 is the correct procedure to use. Having obtained the value of Q_0 which gives the correct value of α at the known temperature, it is, of course, of great theoretical interest to see whether this value agrees with equation (2), but the use of equation (2) to obtain Q_0 appears undesirable.

§ 6. ANOMALOUS THERMAL EFFECTS

The principal danger in using Grüneisen's relation for calculating changes in volume is that anomalous thermal effects may occur which are not related to volume changes by means of equation (1). In some cases (e.g. ammonium chloride, germanium) definite peaks appear on the temperature/specific heat curves, and these can readily be detected. If, however, the effect is less pronounced, the specific-heat curve may appear normal, and the abnormality can only be detected when the expansion is studied. At present the experimental evidence on this point is confusing. In the case of lithium, the work of Simon and Bergmann (1930) suggests that the expansion and specific heats in the range 95 to 273°K . do not obey equation (1), although the expansion data would satisfy this equation if a fictitious characteristic temperature were assumed. The differences involved are not very great, and would be covered by errors of the order 5% in the mean coefficients of linear expansion. Further work appears desirable, since the method used was one in which a rod of metal 30 cm. long was stood vertically in a copper tube. At temperatures above 0°C . the metal was so soft as to undergo deformation, and it is not certain whether the observed coefficients of expansion refer to partly strained material. It is also to be noted that the data of Simon and Bergmann for lithium were accompanied by results for copper which it was suggested showed the same effect, although this was not confirmed by the work of Adensstedt (1936), whilst the detailed work described above suggests strongly that in the range -253 to 0°C . the

thermal expansion and specific heat agree excellently with Gruneisen's equation. It appears, therefore, that much further work is required before these suggested small anomalies can be regarded as established.*

ACKNOWLEDGMENTS

The author must express his thanks to Professor C. N. Hinshelwood, F.R.S., for laboratory accommodation, and many other facilities for research work. Thanks are also due to Dr. F. Simon, F.R.S., for much helpful discussion.

APPENDIX

Notes on the experimental data

Silver. The lattice spacings above 273° K. were taken by linear interpolation between the two adjacent values given by Hume-Rothery and Reynolds (1938), whose work extended over the range 18 to 943° C. These values agree with measurements of other workers on massive silver. The coefficients of expansion at low temperatures are taken from Landolt's Tables (Ergänzungsband II), and are based on the work of Keesom and Jansen (1927). The specific heats are from Landolt's Tables (Ergänzungsband III), which give specially selected values of C_p and C_v .

Copper. The lattice spacings above 273° K. are from the work of Hume-Rothery and Andrews (1942). At temperatures from 87.5 to 273° K., the most accurate expansion data appear to be those of Nix and MacNair (1941), whose results are in excellent agreement with those given in Landolt's Tables (Ergänzungsband II). Nix and MacNair also give data for massive copper at high temperatures, and these are not in exact agreement with the lattice-spacing data of Hume-Rothery and Andrews. The difference is particularly marked at 573° K., where the measured lattice spacing is 3.6259(5) Å., whilst that calculated from Nix and MacNair's data would be 3.6254. If Nix and MacNair's data were used throughout, a slightly different value of Q_0 would give a better agreement with Grüneisen's relation over the range 0 to 573° K., at the expense of a slightly worse agreement at higher temperatures, but at present it seems that the expansion of massive copper at high temperatures varies from one sample to another, and we have thought it best to use the lattice-spacing data. The thermal expansion in the range -253 to -185° C. is taken from the data in Landolt's Tables (Ergänzungsband II). The different volumes of Landolt's Tables give specific-heat data from many sources, whilst Dockerty (1937) gives detailed tables of C_p and C_v from 28 to 194° K. There is good agreement between

* It must be emphasized that no real conclusion can be drawn from the mere fact that a specific heat/temperature curve can be represented as the sum of a Debye curve, and a correction term which reaches a maximum at some temperature. All Debye curves give a zero specific heat at $T=0$, and the same specific heat at very high temperatures. It follows, therefore, that a true Debye curve can always be represented as the sum of a curve for an appropriate value of θ and a correction term which reaches a maximum in some range. The correction term must be of the correct magnitude before its existence can be used as an indication of an anomalous effect, and it is seldom that the specific heats are known with sufficient accuracy. As pointed out by Simon and Bergmann, further work down to the temperature of liquid hydrogen is required before the phenomenon is fully understood.

the results of different investigators below room temperature. At high temperatures the data of Jaeger, Rosenbohm, and Bottema (1933) have been confirmed by Carpenter and Bryant (1939).

Aluminium. The coefficients of expansion at low temperatures are from the work of Ebert (1928), and the lattice spacings at high temperatures are due to Wilson (1942). As in the case of copper, the expansion of the lattice and of massive metal are not in exact agreement at high temperatures. The specific heats are taken from different volumes of Landolt's Tables.

Iron. Nix and MacNair (1941) investigated the expansion of massive metal in great detail over the range -181.5 to $+700^{\circ}\text{C.}$, whilst Ebert's work extends to -253°C. The specific heats at low temperatures were measured by Simon and Swain (1935). Above room temperature the position is very obscure. Several investigators have claimed that the specific heat of iron shows fluctuations, and G. Naeser (1935) gives a specific heat/temperature curve with a series of small peaks and valleys, although these were not found by H. Klinkhardt (1927); both investigators used very pure metals. For the present purpose we have used a smooth curve based on the results of Klinkhardt and Baerlecken. If a toothed curve really exists, it will not greatly affect the calculations, because some parts will lie above and some below the smooth curve, and on integration the positive and negative areas will tend to cancel.

Magnesium. The lattice spacings at high temperatures are from the work of Hume-Rothery and Raynor (1939), whose data for the range 0 to 200°C. are in good agreement with the work of Goens and Schmid (1936), who measured the coefficients of expansion of large single crystals. The work of the latter investigators extended down to -253°C. , and their results have been accepted below 0°C. For the specific heats the data of Seekamp (1931) and of Williams, Eastman and Young (1924) are in good agreement.

Zinc. The lattice spacings at high temperatures are from the work of Owen and Yates (1934), whilst the thermal expansion of single crystals at low temperatures was investigated by Grüneisen and Goens (1924). The two investigators overlap in the region 0 to 100°C. , and are not in exact agreement. For the specific heats, the results of different investigators as given in the different volumes of Landolt's Tables enable a satisfactory curve to be drawn up to 100°C. , but there is a difference of opinion as to whether abnormal effects exist between 100 and 200°C. and between 330 and 340°C. Jaeger and Poppema (1936) give the mean values of C_p between 0°C. and higher temperatures, and we have based our calculations above 100°C. on their figures, which include any anomalous specific-heat effects.

REFERENCES

- ADENSTEDT, H., 1936. *Ann. Phys., Lpz.*, **26**, 69.
 BAERLECKEN, E. *Dissertation*. This is quoted in Landolt's Tables (Ergänzungsband III).
 CARPENTER, L. C. and BRYANT, A. H., 1939. *J. Sci. Instrum.* **16**, 183.
 DOCKERTY, S. M., 1937. *Canad. J. Res.* **15**, 64.
 EASTMAN, E. D., WILLIAMS, A. M. and YOUNG, T. F., 1924. *J. Amer. Chem. Soc.* **46**, 1178.
 EBERT, H., 1928. *Z. Phys.* **47**, 712.
 GOENS, E. and SCHMID, E., 1936. *Phys. Z.* **37**, 389.
 GRÜNEISEN, E. and GOENS, E., 1924. *Z. Phys.* **29**, 141.

- HUME-ROTHERY, W. and ANDREWS, K. W., 1942. *J. Inst. Metals*, **68**, 19.
 HUME-ROTHERY, W. and RAYNOR, G. V., 1939. *J. Inst. Metals*, **65**, 379.
 HUME-ROTHERY, W. and REYNOLDS, P. W., 1938. *Proc. Roy. Soc. A*, **167**, 25.
 JAEGER, F. M. and POPPEMA, T. J., 1936. *Rev. Trav. Chem.* **55**, 491.
 JAEGER, F. M., ROSENBOHM, E. and BOTTEMA, J. A., 1933. *Rec. Trav. Chim. Pays-Bas*, **52**, 78.
 JETTE, E. R. and FOOTE, F., 1935. *J. Chem. Phys.* **3**, 605.
 KEESOM, W. H. and JANSEN, A. F. J., 1927. *Proc. Acad. Amsterdam*, **30**, 576.
 KLINKHARDT, H., 1927. *Ann. Phys., Lpz.*, **84**, 167.
 NAESER, G., 1935. *Mitt. Kaiser-Wilhelm Inst. Eisenforsch.* **17**, 185.
 NERNST, W. and LINDEMANN, F. A., 1911. *Z. Elektrochem.* **17**, 817.
 NIX, F. C. and MACNAIR, D., 1941. *Phys. Rev.* **60**, 597.
 OWEN, E. A. and YATES, E. L., 1934. *Phil. Mag.* **17**, 113.
 SEEKAMP, H., 1931. *Z. anorg. Chem.* **195**, 345.
 SIMON, F. and BERGMANN, R., 1930. *Z. phys. Chem. B*, **8**, 268.
 SIMON, F. and SWAIN, R. C., 1935. *Z. phys. Chem.* **28**, 189.
 SIMON, E. and VOHSEN, E., 1928. *Z. phys. Chem.* **133**, 165.
 WILSON, A. J., 1942. *Proc. Phys. Soc.* **54**, 488.

ILLUMINANTS FOR COLORIMETRY AND THE COLOURS OF TOTAL RADIATORS

BY H. G. W. HARDING

Communication from the National Physical Laboratory

Lecture given to the Colour Group, 28 September 1944; MS. received 23 February 1945

ABSTRACT. The colours of total radiators are discussed and methods of estimating the colour-temperature to be assigned to a colour not on the locus of the colours of total radiators are outlined. The filter method of calibrating tungsten lamps at the National Physical Laboratory is given, with an indication of the probable accuracy of the calibration.

The history and properties are given for the three standard illuminants A, B and C recommended for colorimetry by the Commission Internationale de l'Eclairage in 1931. Other illuminants, such as the equal-energy illuminant E, are mentioned, and the colours of them are compared with those of daylight.

§ 1. INTRODUCTION

THIS lecture deals first with the colours of total radiators and the estimation of the colour-temperature that is to be given to a colour that is not exactly on the locus of the colours of total radiators. Then we consider the standard illuminants A, B and C that are used for colorimetry, what they are, how they are established, and how permanent they are likely to be. After this, an account of other illuminants that have been proposed will be given, together with indications as to how their colours compare with those of daylight.

To begin, definitions of the terms that will be used frequently might be helpful.

Firstly, what is the spectral distribution of energy from a total radiator? A theoretical investigation shows that this spectral distribution of energy of a total radiator is given by Planck's formula,

$$E_{\lambda} d\lambda = C_1 \lambda^{-5} [\exp(C_2/\lambda\theta) - 1]^{-1} d\lambda,$$

where $E_{\lambda} d\lambda$ is the energy between the wave-lengths $\lambda \pm d\lambda/2$ for a radiator at $\theta^\circ \text{K}$. and C_1 and C_2 are constants. It is generally agreed that experimental results are in accordance with those given by this formula (Brown, 1941; Committee on Colorimetry, 1944).

The value recently given for C_2 (Birge, 1941), the only constant that concerns us here, is $1.4384_8 \pm 0.0003_4$ cm. deg. The values suggested for C_2 have varied considerably. 1.4320 was accepted for the international temperature scale in 1927 (*Travaux et Mémoires*, 1927), 1.4360 was considered for this scale in 1939 (Comité International des Poids et Mesures, 1939), 1.4350 has been used

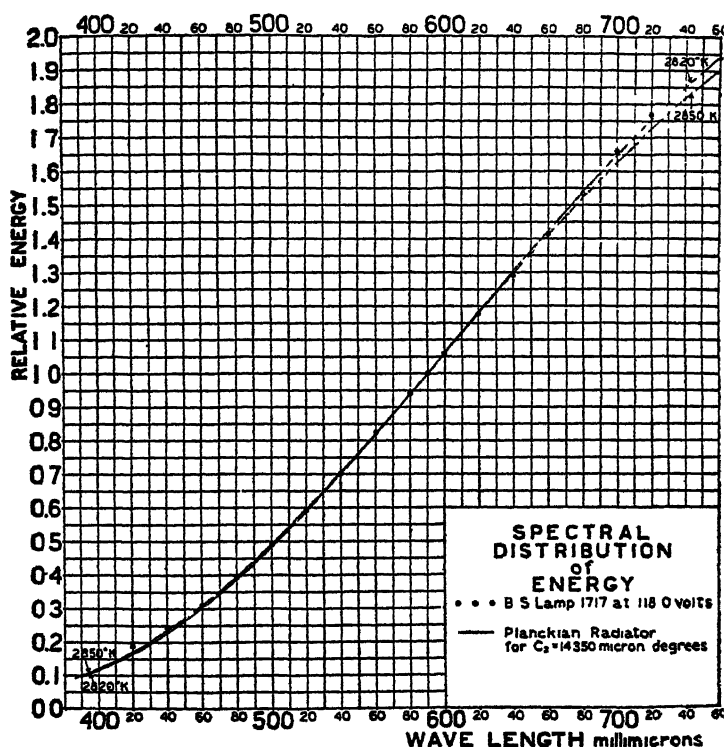


Figure 1 Spectral distribution of energy, B.S. Lamp No. 1717 and Planckian radiator at 2820° and 2850°K

[Reprinted by permission from *Sci. Pap. Nat. Bur. Stand., Wash.*]

extensively for colorimetry since 1931 (Smith and Guild, 1931-2), and no doubt a value higher than 1.4360 will be accepted internationally in the future.

Secondly, what is the spectral distribution of energy from a tungsten-filament lamp over the visible range of the spectrum? Measurements by Coblentz and calculations by Priest show that it is very similar to the distribution calculated from Planck's formula if colour-temperature is substituted for temperature. A graph from Priest's paper (Priest, 1922) is reproduced as figure 1.

Throughout this paper the relative spectral distributions of energy of total radiators and tungsten-filament lamps, for visible radiation, are derived from Planck's formula, and C_2 has the value 1.4350 which was suggested by the Commission Internationale de l'Eclairage, 1931 (Smith and Guild, 1931-2) for the

calculation of the spectral distribution of energy of the C.I.E. standard illuminant A

Thirdly, what is the colour-temperature of a tungsten-filament lamp? By this we mean the temperature of the total radiator which is estimated to have the same colour as the lamp. The international temperature scale (*Travaux et Mémoires*, 1927), which has been used since 1927, is based on the melting point of gold (1336°K.) and Wien's radiation law with a value 1.432 cm. degrees for C_2 .

§2. THE COLOURS OF TOTAL RADIATORS

With these definitions in mind, the colours of total radiators will first be considered. We have found at the National Physical Laboratory that the trichromatic coefficients of the colours of total radiators are often wanted to five-figure accuracy over a large colour-temperature range. Since it takes three or four hours to calculate the colour of a radiator directly from Planck's formula, it was considered worth while to produce tables of the trichromatic coefficients of the colours on the C.I.E. system, so that the colour of any radiator, at any temperature between 1500 and 9000°K. , could be obtained by linear interpolation in a few seconds. In addition, if the tables were smooth to six figures, they could be the basis for other tables that would be required if the value for C_2 were altered or if the colours were to be expressed against the micro-reciprocal degree scale (Priest, 1933), which will be explained later.

Tables were prepared in 1940-1 with a value 1.4350 for C_2 . The energies were calculated from Planck's formula at approximately 500°K. intervals for the temperature range 1500 - $11\,000^{\circ}\text{K.}$, then the colours were calculated from these energy values using the *Condensed Tables for Colour Computation* (Smith, 1934). The trichromatic coefficients so obtained were interpolated for the range 1500 - $10\,900^{\circ}\text{K.}$ The use of Smith's tables about halved the work involved, since the full tables for colour calculations as recommended by the C.I.E. have entries at seventy-nine wave-lengths and Smith's tables entries at only forty wave-lengths. If only half the entries of the C.I.E. tables, that is at 0.38 - 0.01 - $0.77\text{ }\mu$, had been used to calculate the colours from the spectral-energy distributions, errors of 0.0001 to 0.0002 in the trichromatic coefficients would have been expected, whilst with the condensed tables the errors should not exceed 0.00001 .

When in 1941 Birge suggested a value 1.4384_8 for C_2 , the tables were revised to refer to this value by reading from the 1.4350 tables the trichromatic coefficients for the temperatures $\theta_{1.4350} \times 1.4350 \times (1.4384_8)^{-1}$ over the range 1500 - $10\,900^{\circ}\text{K.}$, then interpolating these values to tenths. These tables have been published (Harding, 1944 a). The validity of this procedure can be seen if the value C_2/θ of Planck's formula is considered. If the value of C_2 is changed, the locus of the colours of total radiators will not be altered on the colour chart, but the temperature corresponding to a particular colour on the locus will alter. For example, if $C_2 = 1.4384_8$ and $\theta = 2000^{\circ}\text{K.}$, then $C_2/\theta = 1.4384_8/2000$. If C_2 had been taken as 1.4350 , the temperature given to the radiator having the same colour would be $2000 \times (1.4350) \times (1.4384_8)^{-1} = 1995.2^{\circ}\text{K.}$, because $C_2/\theta = 1.4384_8/2000 = 1.4350/2000 \times (1.4350) \times (1.4384_8)^{-1}$.

The locus of the colours of the radiators is shown in figure 2, where it is

interesting to note that the locus can be represented very closely by a parabola. A better parabola than the one indicated can probably be found, but since the actual temperatures cannot be located on the curve, the only value that the formula has is to serve as a guide to the shape of the curve on the colour chart. The "mired" or micro-reciprocal-degree scale is marked on the locus, and it can be seen that the mired scale is more evenly spaced than the colour-temperature

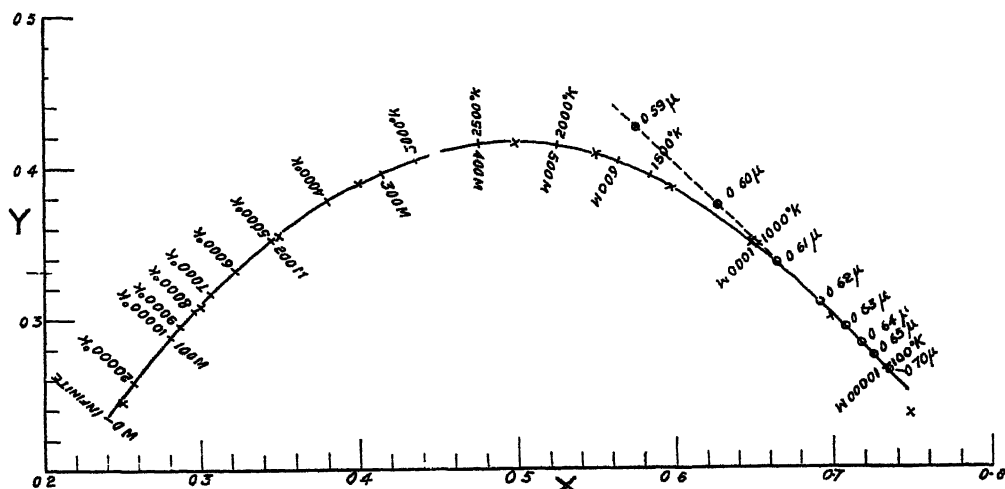


Figure 2 Equation to colour locus of total radiators is approximately $(x-0.496)^2 = -0.358(y-0.415)$, indicated by x

Spectrum colour locus --○--○--
 Total radiator colour locus ---|---|---

scale over the range of temperatures from about 2000° K. to infinite temperature. A mired is a million divided by the colour-temperature in °K., and serves a useful purpose in colour-temperature calculations since the use of it generally avoids calculating reciprocals. Tables giving the trichromatic coefficients for total radiators for the mired range 0-1-660 mireds ($C_2 = 1\,43848$) have been prepared at the Laboratory also.

§ 3. COLOURS NOT ON LOCUS FOR TOTAL RADIATORS

The next problem is that of the colour-temperature to be given to a colour which is not on the locus of the colours of total radiators, and an example will explain why this is so important.

Suppose that a filter is being used for colour-temperature measurements, and that this filter is placed in front of a standard lamp in order to colour-match it with another lamp at a different colour-temperature. It may not be possible to get the photometer fields to be exactly the same colour by altering the lamp voltages because, owing to the properties of the filter, one photometer field is always slightly, but perceptibly, greener than the other. When different observers make colour-matches under these conditions, it is most likely that their settings will not agree, and that one observer would be quite sure, on looking at the photometer fields, that he would never make the setting that another observer would make. When this sort of thing happens, the general procedure

is to take the mean reading of a number of observers, but since a few preliminary measurements will most likely explain the disagreement and avoid observer differences in the final result, they are well worth some consideration.

The mechanism of colour-matching may be illustrated by figure 3. If two lamps L_1 and L_2 are colour-matched at a temperature θ represented by the point P on the locus of the colours of total radiators RS, then a very pale green filter is put in front of L_2 , so that the colour of L_1 is represented by P and that of L_2 by C, an observer will see the colours represented by P and C. To try to restore the colour-match the colour-temperature of L_1 may be altered to correspond to the point B (L_1, θ_1) on the locus RS. The observer can then

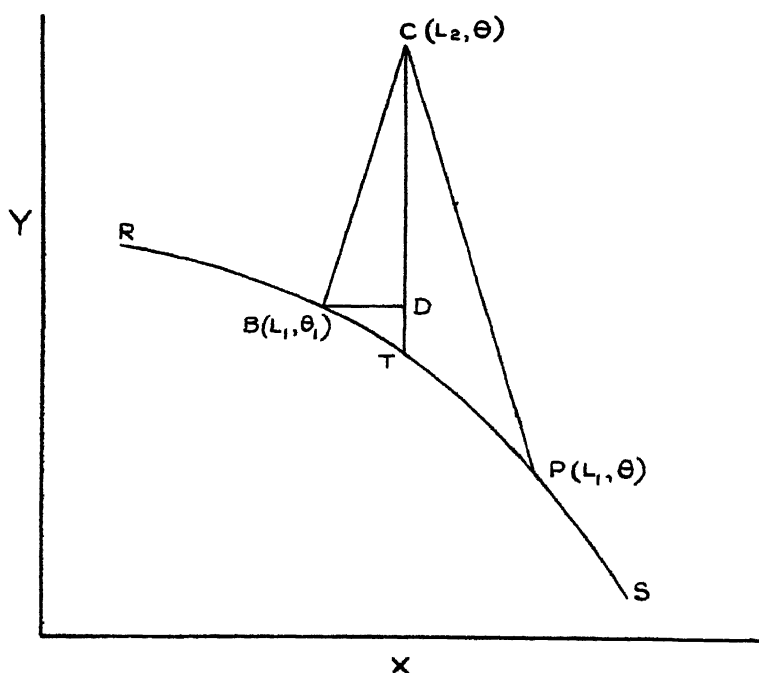


Figure 3. Colour-temperature estimation for colours not on the locus of the colours of total radiators.

be said to estimate colours off the locus with those on the locus along the line CB. If CD is drawn through C parallel to the Y-axis to meet a line through B parallel to the X-axis at D, and to meet the locus at T, the angle DCB can be conveniently defined by giving the value of BD if $CD = 0.001$.

The greenness of the off-locus colour C with respect to the locus RS can be expressed by the value of CT, which is subsequently called Δy . Δy is the difference between the Y-coefficient of the colour represented by C and the Y-coefficient of the colour of the total radiator which has the same X-coefficient. If $\Delta y = 0.001$ and the two colours are viewed with a Lummer-Brodhun photometer having a ten-degree field, there will be an observable colour difference between the two parts of the field, one field always appearing slightly greener than the other however carefully the colour-temperatures of the lamps are adjusted to try to get a colour-match.

About 1930, when the R, G and B system of expressing colours was used at the National Physical Laboratory, the method used to fix the line CB was to divide the red coefficient of the off-locus colour by the blue coefficient, then to find the colour of the total radiator which had the same red to blue ratio. The reason for adopting this method was that experiments by Guild and Young indicated that it gave the correct answer provided that a predetermined method of colour-matching was used.

A second method by Judd (Judd, 1936 a), which was based on the Uniform Chromaticity Chart, is much sounder and will be mentioned later.

This problem of the slope of the line CB arose at the National Physical Laboratory when we were measuring the colour-temperatures of lamps with a blue liquid and a yellow glass filter. The results of the colour-temperature measurements using the alternative filters differed by the easily measurable amount of 10°K. , and it seemed that the most probable explanation of the disagreement was that the colour-temperatures assigned to off-locus colours were incorrect.

Donaldson, of the National Physical Laboratory, determined the way in which several observers estimated the nearest colour-temperature which corresponded to a given off-locus colour. He first colour-matched two lamps at a given colour-temperature using a Lummer-Brodhun photometer with a ten-degree field, then inserted a pale green filter (Δy equal to approximately 0.001) in front of one of the lamps and altered the colour-temperature of the other lamp until the best colour-match setting that could be obtained was found. The slope of CB was then found from a knowledge of the colour of the filter, which had been determined spectrophotometrically, and the calibration of the colour-temperatures of the lamps against their voltages. The results of these measurements are shown in figure 4, where the results obtained by the red to blue (or X to Z of the more recent C.I.E. system) ratio method used by Guild and the Judd method are included for comparison purposes. The letters near to each curve identify the observer.

It is seen from this diagram that if $\Delta y = 0.001$, a value which gives a just-perceptible green colour difference, two observers may differ in their settings by about 12°K. at 2500°K. , and a larger difference than this is likely to be expected for most single-component filters which are available for colour-temperature measurements. If, however, the filters have been carefully prepared, Δy need not exceed 0.00004 (Harding, 1944b), so that observer differences, due to colour differences alone, should not exceed 0.5°K.

Owing to the large variations of individual observers in their estimations of the nearest colour-temperature that corresponds to an off-locus colour, we have made it a practice to use the observer's own estimation of the slope of BC, as given in figure 4, whenever the value of Δy is large enough to cause errors exceeding a degree.

To sum up, there are two factors governing the problem of the colour-temperature that shall be given to a colour. Firstly, if the colour is on the locus of the colours of total radiators, then it is only the value which is given to C_2 in Planck's formula to calculate the colours of the total radiators that matters; that is, illuminant A may be expressed as 2848°K. ($C_2 = 1.4350$), or 2842°K. ($C_2 = 1.4320$),

because $C_2/\theta = 1.4350\ 2848 = 1.4320/2842$ Secondly, if the colour is not on the locus of the colours of total radiators, but is near to it, say Δy not greater than 0.010, there is the additional complication that the value of the colour-temperature estimated by various observers will be different and the characteristics of individual observers will have to be considered.

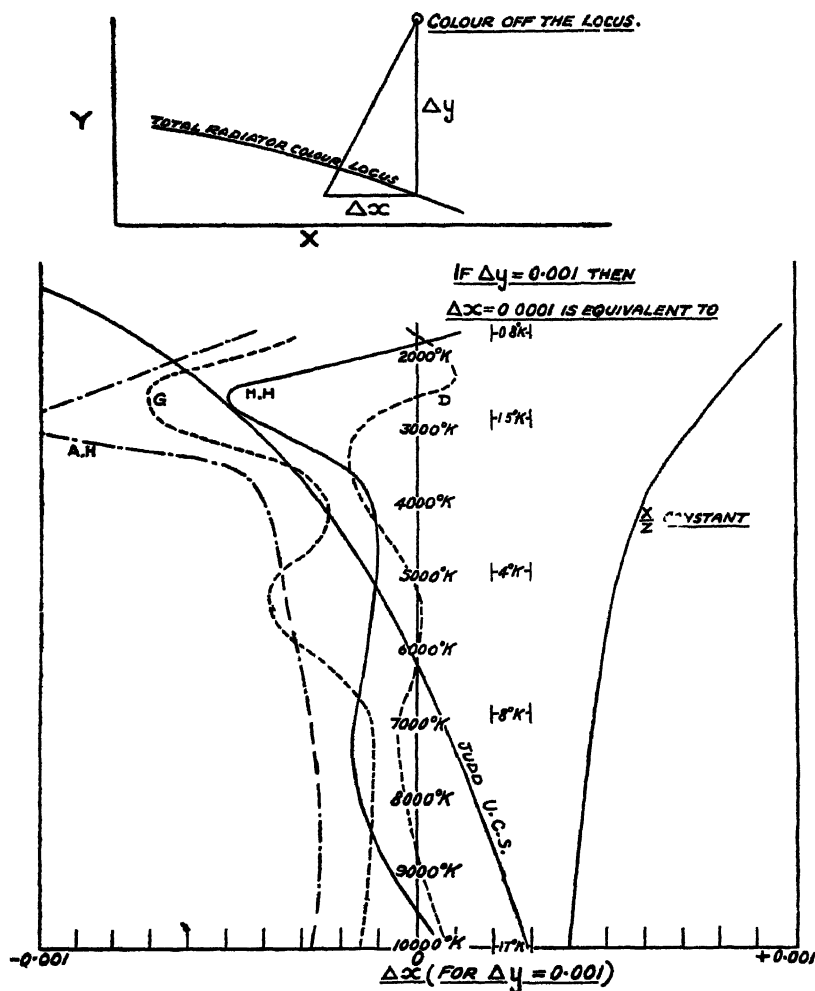


Figure 4. Colour-temperature estimation for colours not on the locus of the colours of total radiators.

When estimations of this kind are made subsequently, it will be assumed that $C_2 = 1.4350$ and that angle DCB is defined by the measurements made by a particular observer at the laboratory.

§ 4. THE STANDARD ILLUMINANTS

We can now go on to the consideration of the first illuminant for colorimetry, standard illuminant A, which is defined by the C.I.E. as "A gas-filled lamp operated at a colour-temperature of 2848°K." "The spectral distribution of

energy from this source may be taken for all colorimetric purposes to be that of a black-body at a temperature of 2848°K . The value assumed for Planck's constant C_2 is 14350 micron degrees" (Smith and Guild, 1931-2).

The value 2848°K . for the colour-temperature of the lamp comes from the United States of America and originated thus: as long ago as 1917, Coblentz of the Bureau of Standards measured the spectral distribution of energy of a five-hundred-watt gas-filled tungsten lamp designated as B.S. Lamp No 1717, operated at 118.0 volts (Priest, 1922). Priest calculated the colour-temperature of this lamp from the spectral distribution of energy in the following way. The values of the products of the visibility function and the spectral distribution of energy of the lamp, and also the products of the visibility function and the spectral distributions of energy of various Planckian radiators, were plotted as ordinates and wave-lengths as abscissae. A value λ_c was calculated for each of these curves, λ_c being the wave-length coordinate of the centre of gravity of a thin template of uniform density bounded by the wave-length axis and the curve.

Priest found that λ_c for the lamp was the same as that for a Planckian (total) radiator at 2848°K ., with an uncertainty of less than 5°K .

A five-hundred-watt lamp was then calibrated against the lamp B.S.1717, and this was sent to the Nela Laboratory. Its colour-temperature, determined by colour-matching with a total radiator, was given by Forsythe in a letter to Priest as 2848°K . (Priest, 1922).

Priest concluded from the good agreement obtained between the colour-temperature derived from Coblentz's isothermal measurements and that measured at the Nela Laboratory by colour-matching with a total radiator, that instead of referring to a particular lamp as a reference standard, a satisfactory standard source would be one closely approximating to the Planckian spectral distribution in the visible spectrum and having a colour-temperature of 2848°K .

This same history for standard illuminant A is also outlined by Wensel, Judd and Roeser (1934), who say that the exactness of the agreement of Priest's and Forsythe's results was apparently accidental and has created the impression that this 2848°K . point on the colour-temperature scale was determined with far more accuracy than was actually the case.

There are two points worth mentioning here. Priest's calculations were made about 1921, which was before the 1924 C.I.E. values for the visibility function and the 1931 C.I.E. trichromatic data were available. It is unlikely that the visibility function that we now use would have affected his results, but it certainly would be interesting to calculate the colour of the lamp from Coblentz's spectral-energy values, using the C.I.E. trichromatic data, to see if the same colour-temperature value of 2848°K . would be obtained.

The calibration of a lamp at 2848°K . is the next problem. By definition a lamp will be operating at 2848°K . if it colour-matches a total radiator which is at a temperature of 2848°K . At the National Physical Laboratory we do not maintain a total radiator in continuous use for routine colour-temperature measurements, but rely on sets of six, sixty-watt, hundred-volt, tungsten vacuum lamps which have been calibrated against a total radiator at several temperatures and for which voltage values have been interpolated at 10°K . intervals over the range 1800-2400 $^{\circ}\text{K}$.

By direct colour-matching, using a ten-degree field Lummer-Brodhun photometer with contrast strips, and taking the mean of eight readings obtained by making two observations for each of the four possible positions of the photometer head, lamps are calibrated from 1800–2400° K. By means of a blue liquid filter (Guild, 1925–6) which is placed in front of the lamp operating at the lower colour-temperature, or a yellow glass filter (Harding, 1944b) in front of the lamp at the higher colour-temperature, the range of the vacuum lamps can be extended to 3000° K., which is about the maximum colour-temperature at which lamps are operated for long periods as standards. To reach the highest colour-temperatures attainable with tungsten-filament lamps, 3644° K. (Priest, 1922), gas-filled projector lamps are calibrated by the filter method against the vacuum standards, then these projector lamps are again used with the filter.

A filter required for these measurements has to be carefully designed. There is only sufficient time here to mention that it must be permanent, very little affected by temperature change, and must give good colour and energy matches between tungsten lamps for the whole range over which standardized lamps might be required. For some purposes a yellow filter is to be preferred, and for others a blue one. Since the best energy and colour matching that we have so

Table 1

Lamp voltage (volts)	Colour temperature (° K.)		
	N.B.S.	N.P.L.	Difference
34.5	2046	2039	7
44.3	2239	2234	5
51.1	2360	2351	9
75.2	2727	2715	12

far been able to get has been with a yellow filter, and since it is an advantage to put the filter in front of the brighter lamp, we often prefer to use a yellow filter.

Before leaving the problems of colour-temperature measurement, it might be worth mentioning what has probably been obvious almost from the beginning of this paper, that is, that a colour-temperature can only be given to a lamp if it exactly colour-matches a total radiator. If the lamp bulbs are slightly coloured, or if the energy from the lamp is not approximately Planckian, then difficulties arise. The colour-temperature scale, however, serves a very useful purpose in defining the operating properties of lamps, and our experience with the filter that we use and the lamps that we have calibrated has been that colour-matches have generally been so good that the observer has not been prepared to say that they are not perfect.

The accuracy of the value of the colour-temperature that is given to a lamp can now be considered. A comparison of measurements made both by the National Bureau of Standards and the National Physical Laboratory in 1934 on a projector lamp are given in table 1.

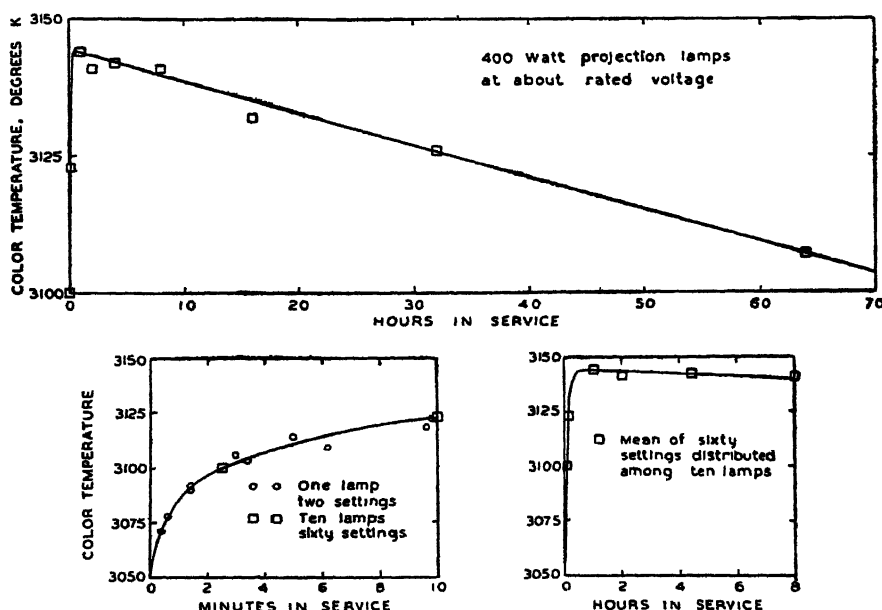
The colour-temperature of a lamp measured with or without a filter can generally be repeated to one or two degrees. Calibrations made with a blue

liquid filter and a yellow glass filter over the range 1950–2848°K. gave results in agreement to 2°K. (Harding, 1944 b).

Next comes the permanency of the calibrations of lamps which are probably similar to those that we use in this country, namely, projector lamps for standard illuminant A and vacuum lamps for colour-temperature standards. The

Table 2

Lamp type	Colour-temperature (°K.)	$\frac{d\theta}{dt}$ (°K./hr.)
Projection	3140	−0.6 ±0.1
	2850	−0.06 ±0.01
	2360	−0.006±0.005
Commercial vacuum	2360	−0.01 ±0.08
Standard vacuum	2360	−0.05 ±0.04



Colours of sunlight and daylight.

Figure 5 Colour temperature of unseasoned lamps as a function of time in service
The colour temperature rises rapidly during seasoning, then declines slowly
[Reprinted by permission from *J. Res. Nat. Bur. Stand., Wash.*]

general behaviour of new projector lamps (Judd, 1936 b) is shown in table 2 and figure 5, both taken from Judd's paper.

The other two standard illuminants for colorimetry are obtained by using liquid filters with standard illuminant A.

Liquid filters were used at the National Physical Laboratory before the C.I.E. 1931 recommendations were made. About 1924 Guild required a

"standard white" illuminant for his measurements of the spectrum colours: these measurements, together with those of Wright, were eventually to form the basis for the 1931 C.I.E. colorimetric system R H Sinden (1923) mentioned a filter which had been developed by Pfund and which, when used with a suitable illuminant, reproduced very closely the spectral distribution of energy

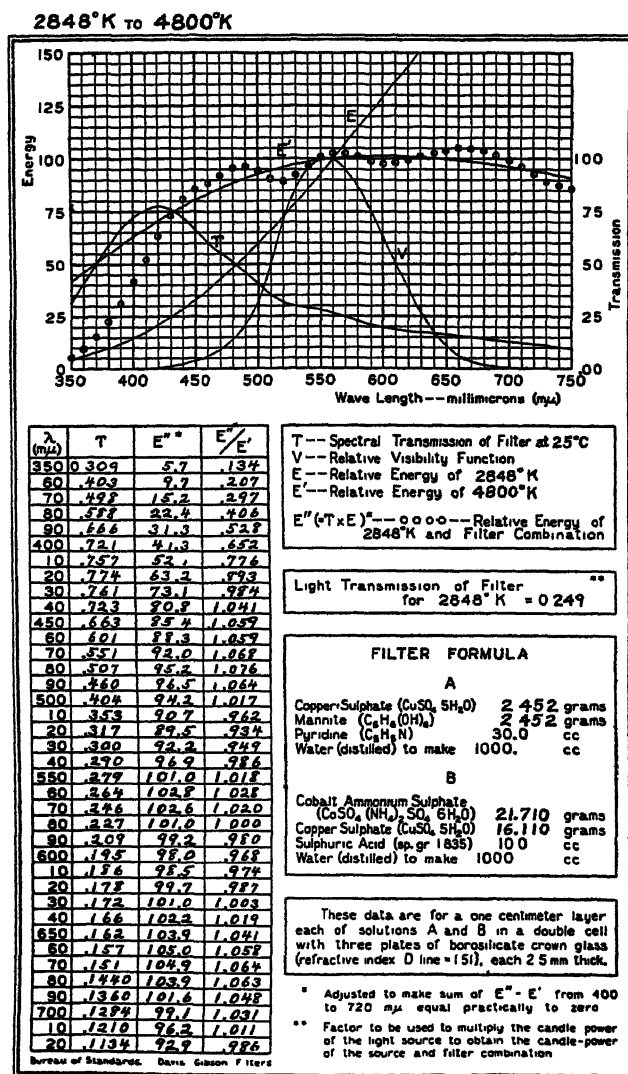


Figure 6.

in sunlight. This filter consisted of a double-compartment cell. One compartment contained an ammoniacal solution of copper sulphate and the other an aqueous solution of the sulphates of copper and cobalt. Guild wrote to Pfund for information about this filter and received a reply stating that Pfund had not yet completed his investigations, so that no information as to the compositions of the solutions was available. Guild's colleague Young therefore

investigated the filters *ab initio* and two filters were developed. One was to raise the colour-temperature of a lamp at 2360° K. to about 2900° K. in order to calibrate a lamp at 2900° K. using the vacuum standards. The other was designed to raise the colour-temperature of a lamp at 2900° K. to about 4800° K. to give a standard "white light" (Guild, 1930-1; Davis and Gibson, 1931 a). These solutions were used at the National Physical Laboratory up to 1931, when the colorimetric specifications were revised, and the C.I.E. recommended the use of a Davis-Gibson filter in conjunction with standard illuminant A (Smith and Guild, 1931-2). The filter chosen was one which had a colour almost the same as that of the N.P.L. white-light filter (Davis and Gibson, 1931 a), so that measurements made before 1931 would still be of value. The description of the filter for standard illuminant B, which with standard illuminant A is

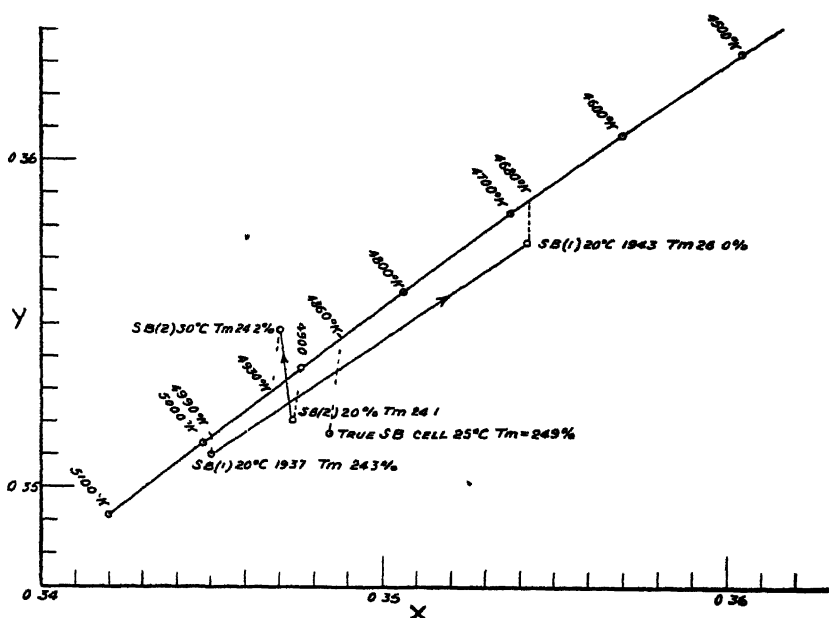


Figure 7. Colours of illuminant B filters for a 2848°K. total radiator as an illuminant.

intended to represent the yellower phases of daylight, is given in figure 6, which is the reproduction of a chart obtained from Davis and Gibson's publication which has just been referred to. The colour of the filter with standard illuminant A as an illuminant is

0.34842X + 0.35161Y + 0.29997Z (Smith and Guild, 1931-2).

the transmission as given by the chart is 24.9 % and the colour-temperature corresponding to the colour is 4860°K. Solutions made at the National Physical Laboratory contained in glass cells (Donaldson, 1933) have colours close to the C.I.E. colour and examples are given in figure 7, where the effects of ageing and temperature change are shown.

It can be seen that the colours of freshly prepared filters are slightly greener than the C.I.E. colour, but the differences are small. Considerable care is

necessary in preparing the solutions if filters made at different times are to repeat their colour characteristics. The procedure that has been adopted is to recrystallize the copper sulphate and cobalt ammonium sulphate, which are of A.R. quality, the crystals are then dried and weighed according to the specification

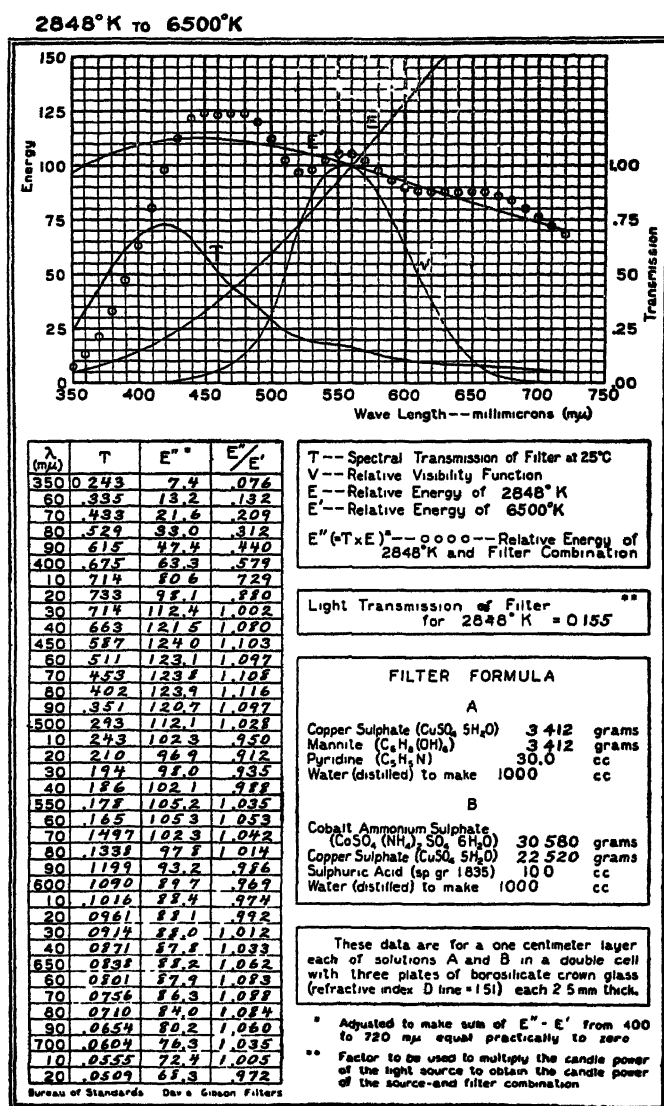


Figure 8

The next illuminant is standard illuminant C, which is standard illuminant A with another Davis-Gibson filter. The illuminant, which is intended to represent the bluer phases of daylight, has not often been used in this country, and we are not in a position to give any of our measurements on its properties. The chart for the filter for illuminant C as received from Davis and Gibson (Davis and Gibson, 1931 b) is reproduced as figure 8.

The colour of the filter with standard illuminant A is

$$0.31006 X + 0.31616 Y + 0.37378 Z \text{ (Smith and Guild, 1931-2),}$$

and the transmission given by the chart is 15.5 %. The colour-temperature corresponding to the colour is 6660° K.

§ 5. OTHER ILLUMINANTS

An illuminant E was mentioned in the *Proceedings of the International Commission on Illumination*, 1939. This is illuminant A with a Davis-Gibson filter (Davis and Gibson, 1934), and the colour of the illuminant is

$$\frac{1}{3} X + \frac{1}{3} Y + \frac{1}{3} Z.$$

The colour of the lamp and filter combination is too far from the total radiator locus for an equivalent colour-temperature to have much meaning, but a probable value is somewhere between 5250–5450° K. The suggested advantages of this

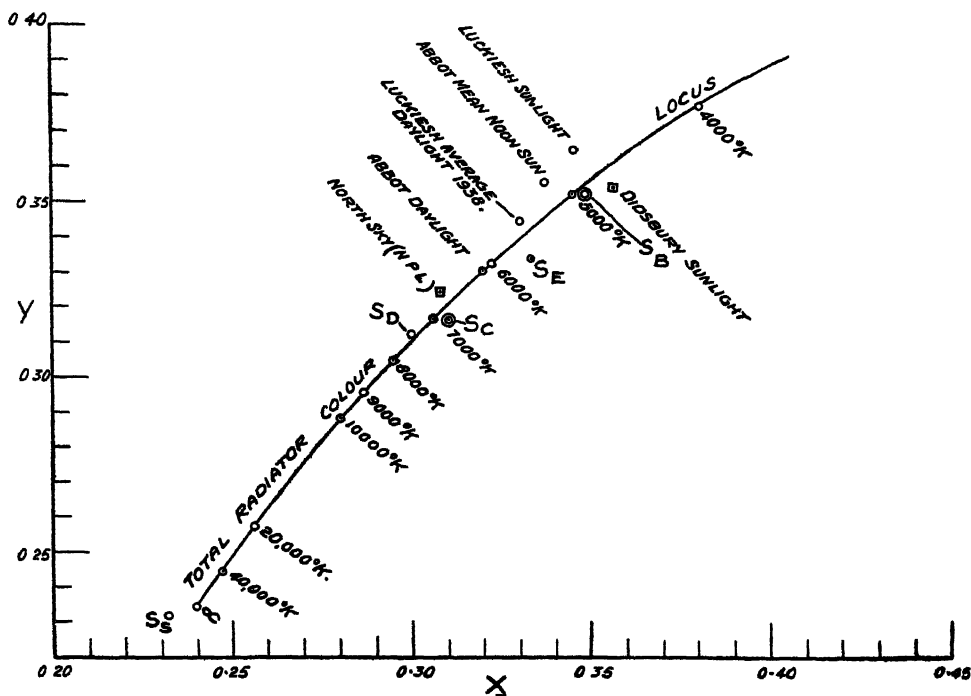


Figure 9.

filter were that it approached daylight more satisfactorily than the standard illuminants B and C, and that many technical and conceptual difficulties could be eliminated by the use of it. These reasons do not seem to be sufficient to justify a change from the illuminants now used, firstly because the colour of the equal-energy illuminant is not closer to the colours of daylight than are the colours of the B and C illuminants, as indicated in figures 9 and 10, and secondly because the technical and conceptual difficulties referred to appear to be of negligible importance.

Illuminants designated D and S (Kelly, Gibson and Nickerson, 1943) have been used in the United States of America. Illuminant D is intended to represent lightly overcast north sky and is stated to have a colour-temperature of 7500°K . It consists of Macbeth (Corning) daylight glass with a lamp at 3000°K . and the colour is given as

$$0.29992X + 0.31201Y + 0.38807Z.$$

This illuminant has been selected for comment because it is made from glass. If melts of glass were reproducible so that the trichromatic coefficients of the colours were reproducible to within 0.001 of a given value, they would be con-

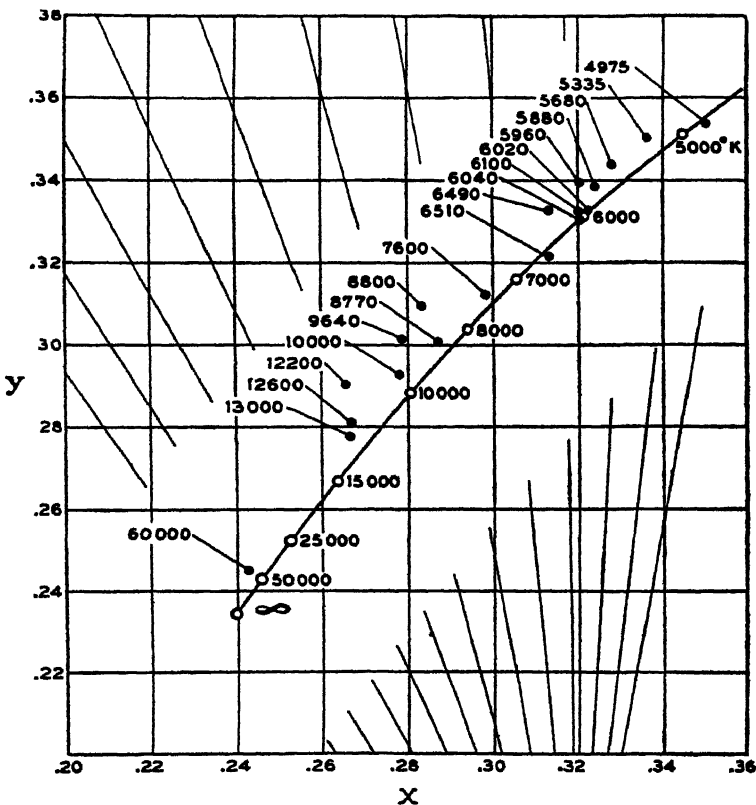


Figure 10 Section of the I.C.I. colorimetry diagram, showing the location of points representing various phases of daylight with respect to the black-body line. The diagonal lines represent various values of "micro-reciprocal degrees" (mireds)

[Reprinted from, and with acknowledgments to, *J Opt. Soc Amer*]

sidered useful filters to use for standard illuminants. Our experience, however, has shown that melts of glass obtained over a considerable number of years differ appreciably, and for this reason single-component glass filters are unlikely to displace liquid filters until glass manufacturers have improved the repeatability of their melts.

Illuminant S is intended to represent extremely blue sky (Gibson, 1940). The trichromatic coefficients of its colour are

$$0.23194X + 0.23176Y + 0.53630Z,$$

The positions of the colours of these filters on the colour chart are shown in figure 9. In addition are given the colours of daylight measured in the United States of America. Abbot Daylight (Nickerson, 1941), Luckiesh Sunlight and Average Daylight, 1938 (Moon, 1941), Abbot Mean Noon Sun (British Standard Specification No. 793-1938); also the colours of daylight measured in Britain, for Didsbury Sunlight and North Skylight (British Standard Specifications 793-1938 and 950-1941). Other measurements of the colours of daylight (Taylor and Kerr, 1940) are given in figure 10

§ 6. CONCLUSIONS

To conclude, there are a few remarks on the colours of daylight and the standard illuminants. Taylor (Nickerson, 1941), commenting on the colours of daylight in America, says that at Cleveland the colour-temperature of a clear zenith sky can be $60,000^{\circ}\text{K.}$, a slightly hazy north sky $12,600^{\circ}\text{K.}$, that smoke in the air reduces the colour-temperature still more, and that a completely cloudy or overcast sky gives an energy distribution which corresponds to approximately 6500°K. He states that in half an hour the colour-temperature of daylight from a north sky may vary by several thousand degrees. Our experience at the National Physical Laboratory is that the colour-temperature of north skylight has varied from 5000°K. to $20,000^{\circ}\text{K.}$ In view of these wide variations in the colours of daylight it is difficult to say which of the illuminants is the most satisfactory.

There is one consideration that should be made when filters are used to reproduce daylight. The bluer the filter the lower is its transmission; for instance, the transmission of the illuminant B filter is about 25 % and that of the illuminant C filter about 16 %, and bluer filters have lower transmissions. For most colorimetric work we require to get as much light of the required quality on the specimen as we can without making it hot. If the filter is put between the lamp and the specimen, it gets hot and its colorimetric properties alter; if the filter is put between the eye and the specimen it is the specimen that gets hot, and this also is not desirable. The transmission of the filter should therefore be as high as possible. This means that the colour-temperature of the radiation transmitted by it should be as low as possible consistent with its colour being in the daylight range. It is probably for this reason that illuminant B has been used so much in this country.

ACKNOWLEDGMENT

Some of the work described above has been carried out as part of the general research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory

REFERENCES

- BIRGE, R. T., 1941. *Rep Prog Phys* 8, 131.
BROWN, T. B., 1941. *Foundations of Modern Physics* (New York: J. Wiley).
BUREAU INTERNATIONAL DES POIDS ET MESURES, *Travaux et Mémoires*, 1927, 18 (Septième Conférence Générale), p. 96.
COMITÉ INTERNATIONAL DES POIDS ET MESURES. *Procès Verbaux*, 1939, p. T 37
COMMITTEE ON COLORIMETRY, 1944. *J. Opt. Soc. Amer.* 34, 198.

- DAVIS, R. and GIBSON, K. S., 1931 a. *J. Res. Nat. Bur. Stand., Wash.*, **7**, 794.
 DAVIS, R. and GIBSON, K. S., 1931 b. Bureau of Standards, *Miscellaneous Publication*, No. 114.
 DAVIS, R. and GIBSON, K. S., 1934. *J. Res. Nat. Bur. Stand., Wash.*, **12**, 265.
 DONALDSON, R., 1933. National Physical Laboratory, *Annual Report*, p. 65.
 GIBSON, K. S., 1940. *J. Opt. Soc. Amer.* **30**, 88.
 GUILD, J., 1925-6. *Trans. Opt. Soc.* **27**, 124.
 GUILD, J., 1930-1. *Trans. Opt. Soc.* **32**, 9.
 HARDING, H. G. W., 1944 a. *Proc. Phys. Soc.* **56**, 305.
 HARDING, H. G. W., 1944 b. *Proc. Phys. Soc.* **56**, 21.
 JUDD, D. B., 1936 a. *J. Opt. Soc. Amer.* **26**, 421.
 JUDD, D. B., 1936 b. *J. Res. Nat. Bur. Stand., Wash.*, **17**, 679.
 KELLEY, K. L., GIBSON, K. S. and NICKERSON, D., 1943. *J. Opt. Soc. Amer.* **33**, 355.
 MOON, P., 1941. *Illum. Engng., N.Y.*, **36**, 322.
 NICKERSON, D., 1941. *Illum. Engng., N.Y.*, **36**, 373.
 PRIEST, I. G., 1922. *Sci. Pap. Nat. Bur. Stand., Wash.*, **18**, 225.
 PRIEST, I. G., 1933. *J. Opt. Soc. Amer.* **23**, 41.
 SINDEN, R. H., 1923. *J. Opt. Soc. Amer.* **7**, 1123.
 SMITH, T., 1934. *Proc. Phys. Soc.* **46**, 372.
 SMITH, T. and GUILD, J., 1931-2. *Trans. Opt. Soc.* **33**, 73.
 TAYLOR, A. H. and KERR, G. P., 1941. *J. Opt. Soc. Amer.* **31**, 8.
 WENDEL, H. J., JUDD, D. B. and ROESER, W. F., 1934. *J. Res. Nat. Bur. Stand., Wash.*, **12**, 529.

ON THE MEASUREMENT OF THE EFFECTIVE AREA OF A SEARCH COIL

BY H. MIKHAIL, M.Sc. AND Y. L. YOUSEF, M.Sc.,
 Physics Department,
 Faculty of Science, Fouad I University, Cairo

MS. received 6 February 1945

ABSTRACT. A condenser null-method for the determination of the effective area of a search coil is described. The underlying principle is to compare the area of the unknown coil with a calculable area, the two coils being used alternately in a powerful, constant magnetic field, of which the region of uniform intensity has been extended by the use of pole-pieces with large area. The flux in each coil as it cuts the field can be exactly neutralized by the charge of a condenser. The ratio of the two areas is given as a ratio of two resistances. An accuracy to four significant figures is possible.

§1. INTRODUCTION

THE effective area of a small search coil containing a large number of entangled turns, such as is used in measuring magnetic fields which are homogeneous only over small volumes, cannot in general be determined with any degree of accuracy by calculation from the dimensions of the coil. The experimental way of calibrating a search coil (Stone, 1934) is by the use of calculable fields. The coil is placed with its plane at right angles to a field H (oersteds) produced by a standard solenoid carrying a current I (amp.),

and the mutual induction M (henrys) between the two is balanced against an inductometer whose primary is joined in series with the solenoid. The inductometer may be so adjusted that, on reversing the current, the flux in the secondary just balances the flux induced in the search coil. Then the effective area A (cm^2) is evidently given by

$$MI = AH/10^8.$$

A number of intermediate steps may be required in passing from the relatively small fields which are accurately calculable to the large fields which it may be desired to measure.

In 1939, Nettleton and Sugden, in connection with susceptibility determinations, described a method consisting in the measurement of the mutual induction between the search coil and a pair of large Helmholtz coils. If the experiment is then repeated with a coil of known area, the ratio of the two areas will be the same as the ratio of the mutual inductances which are measured by balancing against an inductometer.

Obviously, the accuracy in these methods depends on the certainty with which the mutual inductance can be known, whereas the sensitivity depends on the strength of the uniform magnetic field employed. In general, the fields produced by solenoids or by Helmholtz coils are very small compared with the fields produced by an electromagnet.

We shall describe here a sensitive method involving only the comparison of two resistances.

§ 2. APPARATUS AND PROCEDURE

Quite recently, the authors (1944) described, in a work on magnetic-field measurement, how the flux induced in a search coil can be neutralized by the charge of a condenser. The same principle is applied here. The essential features of the experimental arrangement are shown in the figure.

The conical pole-pieces of an electromagnet are unscrewed and replaced by disks or short cylindrical pole-pieces M , M of large cross-sectional area. In our case the diameter of the core of the electromagnet is 7.5 cm. The diameter of the pole-faces of the disks is 11.5 cm., and the thickness is 1.4 cm., the air gap is 6.3 cm. With an exciting current of 8 amp., corresponding to a field of 2685 gauss, the induction in a coil of about 300 cm^2 effective area would produce in the reflecting ballistic galvanometer F a deflection of more than 500 mm. for a circuit resistance of about 2000 ohms.

Referring again to the figure, L is a search coil supported on an ebonite frame J carrying a stout wire D dipping into a mercury cup K to a depth of a few cm. (the optimum depth is easily found by trial). B is a battery of about 10 volts and C is a mica condenser of 1 or 2 μF . capacity. F is a fluxmeter or a long-period galvanometer, shunted by a resistance R which includes the resistance of L and the additional resistance R' . P and P' are two large resistance boxes forming a Rayleigh potentiometer. E is a standard cadmium cell connected with a sensitive galvanometer G across a fixed resistance S , and serves to check the constancy of the potentiometer current. The variable resistance X enables a fixed potential difference to be maintained across S .

When the search coil is suddenly raised from the field, the battery circuit is cut off at K. The condenser, being connected to a closed circuit, is thus discharged, the quantity of charge passing through F being

$$\frac{CVR}{R+F} \text{ microcoulombs,}$$

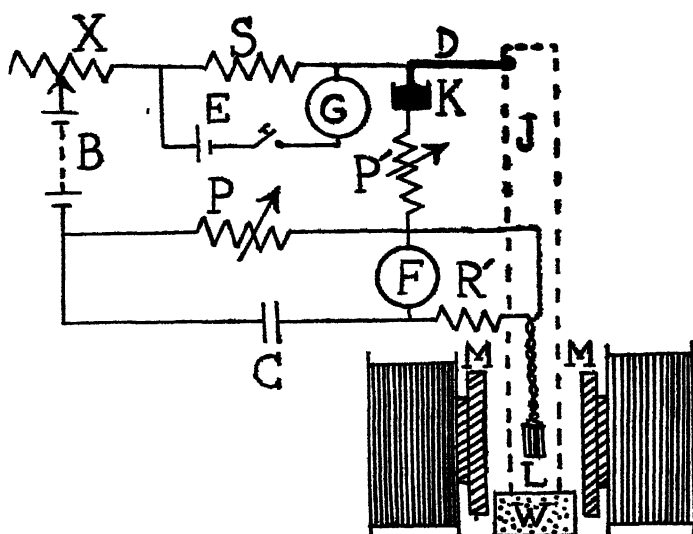
where V is the voltage across P, and F is the resistance of the fluxmeter. Meanwhile, F will receive an opposite impulse

$$\frac{HA}{100(R+F)} \text{ microcoulombs,}$$

due to cutting the field H by the search coil

Exact compensation can be brought about by adjusting the charging voltage or the shunt resistance or both. When this is achieved,

$$A = 100 CVR/H. \quad \dots\dots(1)$$



Under our experimental conditions, the calculated time of discharge of the condenser to $1/10^4$ of the initial charge through P and the shunted fluxmeter is about 0.1 sec. and is of the same order of magnitude as the time of sudden withdrawal of the search coil manually to a distance (about 20 cm. from the gap) where the field is practically nil. The latter time has been determined in a previous investigation with the aid of an Everett-Edgecumbe time-interval meter arranged in conjunction with two mercury contacts, the first being broken at the instant the motion starts, and the second when a distance of 20 cm. has been traversed.

As the times for the opposite impulses are approximately of the same order of magnitude, and can be made to start together, the two effects on F occur almost in synchronism, since the changes at the ends of these intervals are very small. Thus a sharp balance is possible even with a galvanometer whose period is not too long.

It will be seen from equation (1) that for the same values of H , C and V , the area is proportional to the shunt resistance. In other words,

$$A_2/A_1 = R_2/R_1. \quad \dots\dots(2)$$

With the aid of the standard cell, V can easily be maintained constant to 1 in 10^4 .

By fixing the magnetizing current, and reducing its duration to a few minutes so as to avoid excessive electrical heating of the core, it is possible to keep H constant at least to 1 in 1000.

Again, by employing disk pole-pieces, we increase the region of uniform field. In a preliminary experiment with a small exploring coil having a large effective area, it has been found that the field, as measured by the authors' method (*loc. cit.*), is practically uniform, at least up to a radius of 3 cm. from the centre of the gap. Another test of uniformity has been performed by measuring the field with two coils of known area having different linear dimensions.

The two area-coils L_1 and L_2 were constructed on hollow (tubular) ebonite formers turned down to a uniform circular cross-section. Coil L_1 was wound with 36 turns (length 0.9 cm.) of enamelled copper wire of diameter 0.025 cm. on a former of diameter 3.327 cm. The mean radius was thus 1.676 cm. and the effective area of the coil was $A_1 = 317.7 \text{ cm}^2$. Coil L_2 was wound with 56 turns of the same wire on a former of 2.700 cm. diameter. Its length was 1.4 cm. and its area A_2 was 326.5 cm^2 . The windings of the two coils were fixed in position by cellulose dissolved in acetone containing a little ethyl benzoate, the ends bent sharply at right angles and twisted together to reduce their effective area. The coils were then fitted in holes drilled through ebonite frames.

Area measurements were conducted in the following manner. Coil L_1 was held perpendicular to the field and centrally in the gap with the aid of the wax block W . P and P' were then adjusted until the reflecting ballistic galvanometer F showed no deflection on suddenly removing L_1 from the field. Simultaneity of the two impulses could be secured by varying the depth to which the wire D is immersed in the cup K . When exact balance was obtained, the value of the total resistance R_1 shunting F was measured on a good bridge.

Coil L_1 was then replaced by coil L_2 and the experiment repeated. The field current and the potentiometer current were maintained strictly constant, and the previous values of P and C were not changed. Only R_1 was adjusted to R_2 until balance again occurred, and then R_2 was determined.

§ 3. RESULTS

In a typical result, the circuit conditions were:—

$$\begin{array}{lll} B = 12.5 \text{ volts,} & E = 1.0184 \text{ volts,} & V = 5.210 \text{ volts,} \\ S = 1018.4 \text{ ohms,} & P = 5210 \text{ ohms,} & P' = 5900 \text{ ohms,} \\ C = 2.00051 \mu\text{F.} & & \end{array}$$

F was a mirror galvanometer of resistance 881 ohms and period 7.17 sec. A low-resistance galvanometer would have been more suitable, but none was available.

With the area coil L_1 , the shunt resistance $R_1 = 834.0$ ohms.

With the area coil L_2 , the shunt resistance $R_2 = 857.1$ ohms.

Thus,

$$A_2/A_1 = 1.028.$$

By calculation from the dimensions,

$$A_2/A_1 = 1.028.$$

The two results agree to the fourth figure. The agreement may be taken as further indication that the field is homogeneous over a considerable volume.

The unknown search coil L was wound of 100 turns of the same wire as the area coils. It was wound on glass tubing, from which it was afterwards slipped, pressed by thread into a small length and then supported in an ebonite frame. The mean diameter was about 1.26 cm.

With the search coil in the circuit, the shunt resistance satisfying balance under the previous conditions was $R = 348.4$ ohms. The effective area of the search coil L is thus $A = 132.7$ cm².

§ 4. CONCLUSION

The method here described of comparing areas is very satisfactory, and has the additional advantage that the same arrangement which is used for field measurement is employed again, with the modification that the usual conical pole-pieces of the magnet are replaced by ones with larger faces.

REFERENCES

- MIKHAIL, H. and YOUSEF, Y. L., 1944. *Proc. Phys. Soc.* 56, 249.
 NETTLETON, H. and SUGDEN, S., 1939. *Proc. Roy. Soc. A*, 173, 314.
 STONER, E., 1934. *Magnetism and Matter* (London: Methuen), p. 62.

DISCUSSION

on paper by W. E. BALLARD entitled "The formation of metal-sprayed deposits", *Proc. Phys. Soc.* 57, 67 (1945).

Dr. G. A. HANKINS. In presenting the paper, Mr. Ballard did not comment on the mechanical or metallurgical properties of the sprayed metal deposits. In using the process for building-up worn shafts, for example, these properties may be important, since it seems to me there is some possibility that under repeated stresses, as in a shaft, a fatigue crack may be started in the built-up layer, and the stress-concentration at the root of the crack might then be sufficient for the crack to be propagated into the original metal of the shaft and result in a fatigue failure.

The fact that the deposit is able to absorb oil appears to be an asset in the case of a built-up journal bearing. Normally, the engineer endeavours to obtain complete fluid-film lubrication between the journal and a bearing; but in starting or stopping, for example, this oil film breaks down and boundary-lubrication conditions exist. Under these conditions a journal surface which holds a small amount of oil would probably be beneficial.

Mr. F. D. L. NOAKES. I would like to raise three queries with reference to this paper

(1) As it appears that the chief cause of porosity is the boundary layer of particles, which have solidified before they reach the surface to be sprayed, would it not be possible to reduce, if not entirely to eliminate, the porosity by masking?

(2) It did not appear quite clear how it is that although a large proportion of the metallic particles are molten when they strike the sprayed surface, and, therefore, may be assumed to be at quite high temperatures, an inflammable material (or a sensitive substance, such as the human hand) is not in any way affected.

(3) All the samples shown and metals quoted would produce sprayed surfaces which, although they may be matte, nevertheless have considerable metallic lustre. For certain processes various components of radio valves, etc., have to be given an optical black finish. Is there any material which can be sprayed which will produce a surface of this nature?

Mr. L. T. MINCHIN. The real mystery of this process to me is the fact that these minute particles of molten metal escape oxidation. Can Mr. Ballard explain this?

AUTHOR'S reply. In reply to Dr. G. A. Hankins, the paper as presented was solely written with the idea of presenting some theoretical aspects of metal spraying, and, therefore, was not expanded to cover the practical engineering view-point of the process. I have dealt with this in a paper which was presented to the Institution of Automobile Engineers in January 1942, which dealt rather fully with the question of building up crank-shafts. Many thousands of crank-shafts have been treated by the process and have been found to be entirely satisfactory under most arduous conditions.

It is of course possible that a crack started in the built-up layer might cause the crack to be propagated into the original metal, but after a long experience the author has seen only one case of such a failure, and this was in the case of a crank-shaft which had been subjected to a most rigorous engine test with the object of testing to destruction. In commercial practice, as I have said, there has been no such experience, although the possibility cannot be ruled out. The success of this method of reclamation is undoubtedly due to the absorption of oil within the porous steel deposit and the maintenance of an oil film.

In reply to the first question from Mr. F. D. L. Noakes, I feel that, while it would be possible to avoid a good deal of porosity by masking of marginal particles, it is not a possibility in commercial practice. The losses would be so great in proportion that the process would become uneconomic.

With regard to the second question, the particles are extremely small and they are surrounded by a very large volume of gas which is cold compared with the temperature of the particles themselves, therefore the heat transference from the particle is extremely rapid, and this is the reason why sensitive articles may be sprayed. It should, however, be quickly pointed out that while it is possible to spray thin films of cellophane without scorching, it is necessary in such cases either to move the pistol extremely rapidly or to move the cellophane very rapidly; alternatively, the nozzle distance can be increased beyond the usual economic limits. While it is possible to pass one's hand through the spray at working distance, it would not be possible to hold one's hand still in the spray for any appreciable time. The process is considered cold only because of the rapid transference of heat, not because it is really cold.

Finally, in answer to the third question, there is no metal which gives an optical black finish without some form of after-treatment. Iron alloys, nickel chrome and lead give the darkest coatings, and these are dull grey. It would appear therefore that metal spraying is not a solution to this difficulty.

In reply to Mr. L. T. Minchin, it is quite possible to spray the metal magnesium from magnesium wire without undue oxidation. It appears that only two explanations are possible, and one only seems to be of real importance. This is that the speed of the process is so great that the hot particles are not in contact with the hot oxidizing atmosphere long enough to allow of considerable oxidation. The second factor, which may possibly operate in some cases, is that each particle may be covered with a very thin oxide film which serves to protect the particle from further oxidation.

OBITUARY NOTICES

SIR ARTHUR EDDINGTON, O.M., F.R.S.

"LET us now praise famous men". But when the source of fame is untimely stopped, praise comes less readily from our lips than lamentation from our hearts. And less compellingly. It is no accident that our great personal elegies—*Lycidas*, *Adonais*, *In Memoriam*—commemorate those whose renown was unfulfilled. Of Arthur Stanley Eddington it may truly be said that he is dead ere his prime, and hath not left his peer. For though his years were by ordinary standards not few, and his achievements among the greatest which his time can show, yet so vast was the problem towards which his mind instinctively directed itself that the close of 40 years of uninterrupted labours found him still on the hither side of the climax of his thought. We can see none fitted to carry his work to its completion.

Eddington was born on 28 December 1882 at Kendal, of Quaker parentage. He was thus by birth as well as conviction a member of the Society of Friends, and retained throughout his life the essential humility and equilibrium of spirit fostered by the mode of worship of that Society. His early education at Weston-super-Mare was followed by a course and degree in physics at Owens College, Manchester, after which he proceeded to Trinity College, Cambridge, and became Senior Wrangler in 1904. From 1906 to 1913 he was Chief Assistant at the Royal Observatory, Greenwich, where an early inclination towards the major problems of science became definitely oriented, henceforth astronomy was to provide the medium through which he could explore the universal principles belonging to the "nature of the physical world". His first researches, apart from certain routine work (which, however, he never despised, and to the direction of which he gave due attention throughout his life), were concerned with the large-scale movements of the stars, and his work in this field gave such manifest evidence of his quality that in 1913 he was appointed to succeed Sir George Darwin as Plumian Professor of Astronomy and Experimental Philosophy at Cambridge, and in the following year to the additional post of Director of the University Observatory. Here, with his sister, he resided for the remainder of his life.

Eddington's first book, *Stellar Movements and the Structure of the Universe*, was published in 1914. Though the universe there adumbrated is a very modest affair compared with that which it is now usual to imagine, the book remains a classic, both as a masterly summary of the knowledge of the time in this important subject, and as the first intimation to the world in general of the advent of a fresh and powerful mind, master equally of large conceptions and their embodiment in felicitous and intelligible prose. It has a peculiar interest for me because, knowing nothing of the subject, never having heard of the author, and attracted solely by the title, I acquired a copy, found its general meaning within my understanding, and thenceforward saw the sky with new eyes. In *Stellar Movements* appeared those very expressive diagrams which became known somewhat disrespectfully, but not inaptly, as "rabbits", which I believe he was the first to introduce, in which the characteristics of star-streaming appeared in unmistakable form.

It was appropriate that he should sum up his work on stellar movements at this time, for his thoughts were soon to be occupied by two very different and even more fundamental subjects. One of these was indigenous, the natural product of the fertilization by contemplation of the stars of a mind steeped in physical theory. The physical conditions inside a star had been the subject of a few isolated researches, but nothing of obvious significance had been achieved. Eddington, therefore, virtually created a new branch of astronomy when, in 1916, he published his first paper on the equilibrium of stellar matter. The most conspicuous of the new ideas which he introduced was that radiation rather than convection was the chief agent in transferring energy from the interior to the surface of a star, and that radiation pressure played an important part in supporting the weight of the outer layers. From this it followed that the maximum possible mass of a single material body was, in fact, very close to the greatest known stellar mass, and



A. Eddington

PROFESSOR SIR ARTHUR EDDINGTON, O.M., F.R.S
1882-1945

that so long as a star behaved as a perfect gas, its mass and luminosity should stand in a definite relation to one another.

At that time (up to 1924) it was generally believed that giant stars were virtually perfect gases, and that the transition, through contraction, to the imperfect or "liquid" state coincided with the change to the dwarf condition, this was, in fact, the essence of Russell's theory of stellar evolution. Much to his surprise, however, Eddington found that the mass-luminosity relation was obeyed not only by giant but also by dwarf stars—indicating either that his theory was fundamentally wrong or that dwarf stars also were perfect gases. Eddington took the second view, justifying his choice by the consideration that, at the temperatures existing inside the stars, the atoms would be so far stripped of their electrons as to allow abnormally high densities, relatively to which the densities of ordinary dwarf stars were so low that they were effectively perfect gases. The argument was clinched by the existence of "white dwarfs", in which, as Eddington and Russell had previously pointed out, the densities appeared to be impossibly high. The impossible, now become actual, not only proved Eddington to be correct (further confirmation was shortly afterwards provided by the Einstein displacement of the white-dwarf spectrum lines), but also showed the necessity for a completely new theory of stellar evolution.

While the germ of this great idea was beginning to evolve naturally in Eddington's mind, there came from outside a disturbing influence of such magnitude that only the most stable intellect could have preserved itself from confusion. The first account of Einstein's general theory of relativity appeared in 1915, but owing to the war it was not until late in 1916 that, through a series of papers by de Sitter in the *Monthly Notices* of the Royal Astronomical Society, it came to the notice of English scientists. It is, I think, one of the most striking evidences of Eddington's mental power that not merely did he protect the development of his work on stellar interiors from interruption by his inevitably strong reaction to this profound conception, but, grasping the essence and implications of relativity with almost incredible speed, he became at once the leading (for some time the only) English exponent of a theory which, to most of his contemporaries, was so recondite as almost completely to elude apprehension. Nor did he merely hold the mirror up to Einstein. The light of relativity was not reflected in his diffusion of it, but absorbed and re-emitted in characteristic radiations (forbidden radiations, some might consider, but that is an open question). Those who, after a quarter of a century of continuous meditation on the subject, feel that at last they appreciate its general significance have only to turn to Eddington's *Report on the Relativity Theory of Gravitation*—published, we are proud to recall, by this Society—to realize that what they have mastered with so much effort was already well under his control in 1918. So remarkable was this achievement that I once asked him if any aspect of the idea had previously presented itself independently to him. He answered, "No"; Einstein's theory had come as a complete novelty and had changed the direction towards which his general philosophy had been moving. It is tempting to attribute this to a general receptiveness of new ideas, but subsequent events make this hard to sustain. In most matters the groove of his thinking was cut so deep that only with difficulty, if at all, could the thoughts of others enter it. Eddington's philosophy took in the main a pre-determined course, and we must conclude that it was already prepared for the stream of relativity which was to flow along it as readily as though it had cut its own passage. Strange flowers were to grow along the banks, not yet fully to be appreciated, but of that hereafter. It must be remembered that Eddington saw every problem *sub specie aeternitatis*, and to such a mind a principle like that of relativity must have insisted on immediate acceptance or rejection. A decision could not be postponed, for on it depended the light in which every problem was thenceforward to be seen.

The Mathematical Theory of Relativity appeared in 1923. In this book, and particularly in the Introduction, we get the first, and in some respects the clearest, account of the fundamental view of physical science which determined the whole ensuing development of his ideas. It deserves the most careful study, for the manner in which the impact of relativity changed the pre-existing conception of the meaning of physics was markedly different in Eddington and in most of his contemporaries; and herein lies, beyond doubt, the source of the difficulty which other thinkers have had in understanding his later work. Very briefly, the position may be stated thus. In pre-relativity physics an external world was contemplated, having a fixed objective character which it was the object of physics

to discover and describe. Physical measurements were the direct determination of objective features of this world, and as such had an essential independence of the observer. They might, of course, be erroneous, but there were actual correct values to which they approximated more or less closely. Relativity destroyed this view by showing that in all the measurements we commonly made there was an inescapable subjective element, which had to be separated out before one could legitimately draw any conclusions about an external world. The problem was thus posed. How shall we now regard the relation between our measurements and the external world? I think it is correct to say that the majority view—represented, for instance, by the various schools of logical empiricism in philosophy and by Bridgman and other “operationalists” in science—is that measurements exist first of all in their own right, representing only the operations which yield them, and that our picture of the external world must take the form imposed on it by the necessity of integrating those measurements into a rationally coherent system. We start, therefore, not with an external world which we imperfectly observe, but with observations themselves, our description of the external world being our attempt to rationalize them.

Eddington took a view which lay between this and the older view which it displaced. He acknowledged the purely operational definitions of measurements and physical quantities, but posited in addition an external world having certain “conditions” which were related in some undefinable manner to the measurements. When you measured a length you did not, as in the older view, discover the magnitude of an objectively existing length in the external world, you determined a “physical quantity” which was “defined by the series of operations and calculations of which it is the result.” This is pure operationalism, but Eddington added the postulate that the measurement represented in some way a “condition of the world” existing objectively and independently outside. Very different operations, such as measurements of length and parallax, for example, might correspond to the same condition of the world, hence the relation between such conditions and physical quantities was not one of resemblance. We could not, in fact, say anything at all about the relation, except that it existed; the “world” with its “conditions” was essentially mystical and unknowable, all we could do was to deduce its “structure”, never its “essential nature”, and even of that we knew only that it was expressed metaphorically by the mathematical relations found between physical quantities.

This altogether inadequate, but not, I think, false, account of Eddington’s philosophical axioms may perhaps help to explain why his later work did not meet with the understanding which he expected. The inscrutable “conditions of the world” hung like the Old Man of the Sea round the neck of his thought, contributing nothing and serving only to retard its progress and obscure results which, expressed simply and directly in terms of the essential measurements alone, might have commanded understanding and acceptance. This was particularly unfortunate because, from this point onwards, his attention was given more and more to the construction of what, for want of a better term, might be called a general theory of knowledge—a systematic scheme of conceptions which, seen from one side, is theoretical physics and, from another, scientific philosophy. It is impossible to separate Eddington’s general scientific work from his philosophy, and, indeed, in this respect he was well in advance of most of his contemporaries, to whom science is one thing and philosophy another, the one which they do not favour being spoken of with greater or less disrespect. It is therefore all the more regrettable that he who in many ways was better fitted than anyone else to annihilate a false antagonism was forced to express himself in a language which neither party could understand. A series of publications—*The Domain of Physical Science* (in *Science, Religion and Reality*), *The Nature of the Physical World, Science and the Unseen World, The Expanding Universe, New Pathways in Science, Relativity Theory of Protons and Electrons*, and *The Philosophy of Physical Science*—all but one written for the layman as well as the physicist, present the ideas of the theory in its various aspects.

What such a mass of literature, coming from a master of exposition about work of his own creating, has failed to make clear it would be idle to attempt to outline here. Suffice it to say that the theory was an attempt at harmonizing relativity and quantum physics by presenting the characteristics of protons and electrons as characteristics of the universe seen from a particular angle, the smallest things in our map of the structure of the world were equivalent to the largest, in that they shadowed forth the same

"condition of the world". This was not an empty paradox. A mathematical scheme was constructed in which the radius of the universe was expressed in terms of the number and mass of the elementary particles. Some of the relations deduced are certainly impressive, and Eddington was convinced that they were deduced by pure reason, without assistance from knowledge obtained by experiment or observation. He did not, as has sometimes been asserted, claim that all knowledge, even in physics, was obtainable in this way; it was the fundamental laws and constants of nature, not the relatively accidental characteristics of particular objects, that he claimed to be able to derive from the principles of thought alone.

It is natural for those who knew Eddington only through his writings to look upon this work as the product of an illusion, and, indeed, in terms of objective likelihood, it is difficult to do otherwise. Those who knew him personally, however, will, I think, agree that the matter has another aspect. The idea of the "great man", different in kind and not merely in degree from the general run of humanity, is (no doubt rightly) out of fashion, yet it was impossible not to feel that Eddington was gifted with an exceptional insight into the fundamentals of scientific thinking, and had an intuitive apprehension of things which, when they can find expression in generally acceptable terms, will be seen to be both true and vitally important. There was an air of conviction about him which bore the stamp of genuineness. I would venture the opinion that his failure to convince others was due neither to the falsity of his ideas nor to any defect in his expression of them in terms satisfactory to himself, but to that mental peculiarity to which I have already alluded, which prevented him from understanding the way of thinking common to most of his colleagues. In matters of common agreement his complete mastery of his subject was unmistakable. Here, you felt, on reading his account of something which you already understood, is the perfect exposition; no one can fail to understand and acknowledge the inevitability of this reasoning. But when you approached a paper on his own theory, though you recognized the same master at work and found no sign of hesitation or uncertainty in the march towards the goal, you followed in blank incomprehension, soon losing all power to be surprised even if at length you came to something patently absurd.

The situation was exasperating. The conscientious thinker could not ignore what he felt to be the duty of reading the riddle, and enquiry served only to show more clearly than ever the impenetrable barrier of incompatibility between Eddington's fundamental axioms and his own. To take a single example, after struggling with what must have been one of his last and most fundamental attempts to explain his ideas—"The Evaluation of the Cosmical Number" (*Proc. Camb. Phil. Soc.* 40, pt. 1, 37, 1944)—I wrote him, not many weeks before his death, asking for light on a number of points, both general and particular. For instance, "The use of the terms 'measure', 'measurable', etc.", I wrote, "in the 'prologue' to physics inevitably suggests that the things represented thereby are later to be identified with the things so named in ordinary physics, and one cannot help thinking of them in that way. This seems to be encouraged by the example of distance on p. 38. Yet later I read 'The basal measure . . . is the energy tensor', and 'a measurable is one kind of particle'. There is a sense in which I can agree that a measurement of distance involves four entities, but no sense at all in which I can say the same of a kind of particle". The reply to this was: "You are a bold man to set any limits to what a 'particle' may mean since the dear old classical particle became obsolete." This liberty which he assumed of tacitly changing the meaning of words with the progress of ideas is, indeed, a clue to his thought, and those who cannot think the dear old classical particle will let itself be snuffed out by an article may be helped by the realization that any of his apparently incomprehensible remarks about familiar things may be unrecognized re-definitions. But it is far from being a complete clue, and much further labour is necessary before the ordinary reader can reach the point of view from which Eddington's account of his ideas really expresses their meaning.

He himself was as puzzled as others by the situation. "I am interested in your detailed criticism", he went on, "because I am continually trying to find out why people find the procedure obscure. But I would point out that even Einstein was once considered obscure, and hundreds of people have thought it necessary to explain him. I cannot seriously believe that I ever attain the obscurity that Dirac does. But in the case of Einstein and Dirac people have thought it worth while to penetrate the obscurity. I believe they will understand me all right when they realize they have got to do so—and

when it becomes the fashion 'to explain Eddington'." From this view I would not care to dissent.

I do not think I am deceived in discerning in his later work a tendency towards a point of view which I at least might find it possible to share with him and from which his ideas might appear in intelligible form. The later part of *The Philosophy of Physical Science* (his last published book) shows for the first time a willingness to question the axiomatic basis of his philosophy, and gives some indication that he might ultimately have brought it into conformity with that of others. It is probably this development that Professor Temple had in mind when, in a passage with which I wholly disagree, he wrote (*Observatory*, 66, 9 (1945)) that Eddington's "professions of subjectivism or solipsism were semi-jocular exaggerations." They are neither properly denoted by the term "solipsism" nor, I am convinced, anything but expressions of fundamental convictions. On such matters Eddington could not be trivial. He might be profoundly wrong, but he would inevitably be profound. His playfulness found sufficient scope in his method of exposition, where, indeed, though always delightful, it was sometimes an obstacle rather than a help to understanding. It is picturesque to imagine the electron measuring up the local radius of curvature of space-time and adjusting its dimensions accordingly, but the metaphor does not bring to all minds an immediate understanding of the underlying physical relation. Nevertheless, the relation was an essential part of his philosophy, and in the matters to which Professor Temple refers, the truth towards which, as it seems to me, he was unconsciously moving was simply that since the ultimate evidence for any assertion at all one makes about the world must lie in his own experience (even if, to use ordinary language, it rests on someone else's reasoning or experience, one is able to adduce that only on the evidence of his own experience of hearing or reading of it), a systematic deduction from evidence to conclusion must necessarily begin with the experience one has one's self. In particular, the assertion of the "existence of other people" must be established on this necessary basis if risk of self-contradiction is to be avoided. Eddington had long since realized the necessity for this, but had not taken the final step of yielding to it completely. In *The Nature of the Physical World* (1928) he wrote "The only subject presented to me for study is the content of my consciousness". In *The Philosophy of Physical Science* (1939) he returned to the subject, and felt the necessity of reconciling this view with his references to "other people", but did not get further than arbitrarily postulating that other people must be supposed to exist. Possibly his reluctance to accept the inevitable implication of his doctrine arose from the fear that an attempt to formulate the meaning of the phrase "the existence of other people" in terms of the experiences of his own which gave it legitimate significance might degenerate into, or be regarded as equivalent to, a denial that other people "exist" in the absurd sense in which that doctrine is generally denoted by the name "solipsism". I am convinced, however, that, had he lived, he would have removed this defect in his reasoning, and in 1940 I wrote (*Observatory*, 63, 25) "I venture to prophesy that, if he is granted the years an astronomer has the right to expect, he will take this step." Those years, alas! were not allowed him.

As a man Eddington had a strongly attractive personality with which, paradoxically, it was difficult to come into close relationship. Tall, well-built, handsome in youth and an imposing figure always, he would draw a second look in any company. His interests were many, and not confined to the realms of pure reason. He had a wide knowledge and keen appreciation of English literature, but could yet enjoy a well-conceived detective story without being either unaware of or unduly repelled by its literary mediocrity. He was a crossword puzzle addict, and enjoyed a game of golf, though his favourite way of spending a holiday was to mount his bicycle and disappear as completely as possible from human ken. This practice, like the most obvious of his characteristics—an extreme shyness—might easily but wrongly be taken to denote a natural shrinking from human intercourse. On the contrary, he appeared to desire fellowship with others who shared any of his numerous interests—a quality nowhere more in evidence than at the dinners of the Royal Astronomical Society Club, of which he was a member for many years and, at the time of his death, Vice-President and a "double-centurion". His reserve appears rather to have originated in the natural difficulty which he always had in bringing his thoughts into line with those of others, coupled with a slowness of expression which was perhaps not unrelated to, though not entirely a consequence of, a scrupulous care for ultimate exactness. Of the first point, the peculiarity of his outlook on fundamenta

physical problems, already alluded to, is an outstanding example ; it remains only to add that, unlike most shy people, he had little self-consciousness, but was strongly self-centred. If he reviewed a book (which he rarely did, knowing his natural unfitness for the work) he would tell the reader, of whose expectations he appeared to be quite unaware, little or nothing of the author's aims and achievements, but would deal only with the points, no matter how subsidiary, which related to his own work and interests. This was not *because* they related to his own work, the work rather had become his own because he thought it of major importance, and he appears really to have believed that he was giving a proper delineation of the book by restricting attention to it.

His slowness of expression revealed itself most prominently in unprepared discussions and controversies. Here he would show to such poor advantage that a stranger might have been pardoned for imagining that he had not previously given any thought to the subject. He would begin a sentence, pause, start another, return to the first, the whole process being accompanied by characteristic slight and unrevealing gestures of the hands and lips, and finally sit down abruptly without even reaching the anti-climax which was well in sight, leaving the other controversialist to wonder whether the discussion was ended, and, if not, whether he could do anything more useful than repeat his previous point. When the written account of the discussion appeared, however, the chances were that Eddington's case would appear complete and final. It was the same in private conversation. Unless one knew him well, a talk was apt soon to peter out in an awkward silence which it was difficult to terminate. The point of discussion having been dealt with (not necessarily settled), his interlocutor had the feeling that a mental chasm which had momentarily been bridged had opened up again. Eddington, too mistakenly kind-hearted to walk abruptly away, and yet finding nothing to say, would lapse into his own meditations, while his companion would stand uncomfortably waiting for the occurrence of some process of physical disconnection for which no natural mechanism existed. A distinguished foreign astronomer, familiar with his writings and expecting, on meeting him for the first time, to hear a ready elucidation of his problems, said afterwards in a tone of bewilderment: "I was never more surprised. He can say 'Yes', and he can say 'No', and that is all that he can say!"

Eddington had a keen sense of humour, and a subtle though not a ready wit. Here, again, time was necessary for the manifestation of his qualities. An occasion in 1928 comes to mind on which, running out of matches for his indispensable pipe, he entered a shop in a small German town and found that the necessary word had escaped him. A period of ineffective murmuring at length culminated in the word "Lucifer!" uttered explosively at the startled and uncomprehending girl behind the counter. Only later did the master of expression get his meaning across by the act of striking on an imaginary box. Yet in retrospect he could enjoy the situation as well as another.

I have said little of Eddington's astronomical work in the relatively narrow sense, outstandingly important though it is, for that comes more fittingly within the province of another Society. To us he was a physicist in the most fundamental and comprehensive sense of the word, whose Presidency of our Society is among our most cherished traditions. Honours were showered upon him, which also will be recorded elsewhere. He received them with pleasure and gratitude and with no affectation, knowing well that they were but the guinea stamp, and conscious, without false conceit, that they were impressed on gold. This notice makes no attempt at a complete presentation of this many-sided genius. I have tried merely to indicate what seem to me to be the most significant elements of his work, and at the same time in some measure to preserve for those who knew him, and create for those who did not, a partial picture of one of the most remarkable men whom it has been my good fortune to know.

HERBERT DINGLE.

CHARLES GLOVER BARKLA, F.R.S., NOBEL LAUREATE

THE name of Barkla will always be associated with the K and L characteristic X-radiations which his brilliant series of researches revealed before the nuclear theory of atomic structure had been enunciated, and while the discriminating feature of a beam of X-rays was still its penetrating power.

Born in Widnes on 7 June 1877, Charles Glover Barkla was the son of John Martin Barkla, secretary to the Atlas Chemical Company of that town. He was educated at the Liverpool Institute and at University College, Liverpool, and graduated in 1898, obtaining first-class honours in physics. In the following year he was awarded a research scholarship by the Royal Commissioners for the Exhibition of 1851 and proceeded to Cambridge, entering Trinity College as an "advanced student" admitted to a course of research. Barkla's early work at the Cavendish Laboratory was an investigation of the velocity of electric waves along wires, a research in which he used a magnetic detector of the type designed by Rutherford. An account of this work is published in the *Philosophical Magazine* for 1901. During his third, and last, year at Cambridge, Barkla began the comprehensive series of experiments with X-radiations which brought him recognition as an experimentalist of the highest rank.

It had been observed by Röntgen, who seven years earlier had discovered X-rays, that a gas exposed to these rays became a source of secondary radiation. This radiation was the subject of Barkla's experiments. At an early stage in his investigations he discovered that the secondary rays were of two kinds: one kind, which was obtained when the primary beam traversed gases or solid elements of low atomic weight, was found to be similar in quality to the primary radiation, while the other kind, obtained most strongly when elements of high atomic weight were irradiated, was always found to be more absorbable than the primary radiation which gave rise to it. Barkla showed that the former was produced by scattering of the primary beam, the scattering centres being the electrons in the atoms of the irradiated substance. Using the theory of scattering developed by J. J. Thomson, which assumed that X-rays consisted of a succession of electromagnetic pulses, Barkla calculated the number of electrons in the atoms of the various light elements on which he experimented. His results were corroborated later by Rutherford's measurement of the nuclear charges of these atoms. Barkla investigated also the angular distribution of the scattered radiation, and obtained results which gave further support to Thomson's theory of the nature of X-radiation. This stage of his work culminated in the establishment of the fact that the secondary radiation from carbon, scattered in a direction perpendicular to that of the primary beam, was plane polarized, a result of fundamental importance in connexion with the conflicting views then held as to the nature of X-rays.

The other kind of secondary radiation—that which was more absorbable than the primary—was found to be characteristic of the irradiated substance and, moreover, could only be excited when the incident radiation had a penetrating power exceeding a certain minimum, peculiar to the substance. Its production was accompanied by a selective absorption of the primary, and, in contrast to the scattered rays, this characteristic or "fluorescent" radiation—as Barkla named it—appeared at first to be homogeneous, even when the primary beam exciting it comprised radiations of widely different powers of penetration. Barkla subsequently showed that the characteristic radiation could be resolved into two components of very different penetrating powers, and further, that on irradiating elements of successively increasing atomic weights, the resulting characteristic radiations were increasingly more penetrating. Of these two component radiations the more penetrating, or K radiation, proved so homogeneous in regard to absorbability that it was regarded as constituting a single spectrum line, whereas the less penetrating component—or L radiation—showed heterogeneity, which suggested that it consisted of more than one line.

The absence from the characteristic radiation of any evidence of polarization, its uniform distribution around the radiating substance, and the manner of its variation in intensity with variation of quality of the primary beam, all of which phenomena Barkla discovered in the course of his investigations, showed that it was produced by a process very different from that which gives rise to the scattered rays. Barkla's first attempts to explain how the characteristic radiation originated were necessarily very tentative. Experimental results indicated that a much more important part must be assigned to processes taking place within the atoms of the absorbing substance than had been necessary in accounting for the scattering. As Thomson's theory had proved so successful in interpreting the scattered radiation, an attempt to extend it to cover also the characteristic type was the most natural line of approach. Since the radiation was characteristic of the irradiated element, and was more easily absorbed than the radiation giving rise to it, it

could not originate simply in the response of a free electron to the intense electric field of the primary pulse. The fact that it was emitted most strongly by elements of high atomic weight suggested to Barkla that an explanation might be found by considering the disturbing forces exerted by neighbouring charged particles on an electron set in vibration by an incident electric pulse. Such disturbances, which would be most marked in the case of elements of high atomic weight, on account of the greater number of electrons in their atoms, might result, Barkla thought, in the emission of a pulse more complex, and hence more easily absorbed, than the primary pulse.

Barkla continued to discuss his results in terms of ether pulses until 1913, when the diffraction of X-rays by crystals demonstrated the essential similarity of X-rays and ordinary light, and confirmed the conception of X-ray wave-lengths. At about this time, ideas concerning the structure of atoms were being considerably clarified. Rutherford established the existence of a small massive positively-charged nucleus, Bohr published his theory of the arrangement of extra-nuclear electrons, and Moseley's photographs of X-ray spectra, showing the progressive shift of corresponding lines towards shorter wave-lengths as the atomic weight of the emitting element increased, indicated that the inner shells of electrons were responsible for the emission of the characteristic rays. Attempts to account for these in terms of disturbances within a Rutherford-Bohr atom were made by Thomson, Bohr, and Moseley, but Barkla propounded a more satisfactory interpretation. His views were published in 1915 and elaborated in the Bakerian lecture which he delivered to the Royal Society in the following year. His experiments had convinced him that, for each electron emitted directly from the K shell of an atom by absorption of energy from the primary beam, there was emitted also one quantum of characteristic K radiation, possibly accompanied by quanta of radiations of smaller frequencies. He had shown, further, that the emitted electron took no part in the production of the characteristic radiation, and he therefore concluded that the latter originated in some process taking place within the ionized atom as it regained the normal state. From a consideration of the energies of the characteristic and of the corpuscular radiations, together with the magnitude of the absorption of energy from the primary beam, Barkla showed that this readjustment took place in successive stages. The first process, which was accompanied by the emission of K radiation, consisted in the replacement of the emitted K electron by one from the L shell; the filling of the gap thus created in the L shell by an electron from the M shell caused the emission of a quantum of L radiation; and so the process continued through successive shells until, finally, a free electron was captured into the outermost shell of the atom. For the discovery of the characteristic radiation and this explanation of its origin Barkla received the Hughes medal of the Royal Society in 1917, and was awarded the Nobel prize for physics for the same year.

The long series of investigations which Barkla began in Cambridge was continued by him in three different universities. He left Cambridge in 1902 on his election to the Oliver Lodge fellowship of the University of Liverpool, and during the years 1905-9 he was successively demonstrator, assistant lecturer and special lecturer in the physics department of that university. In 1909 he succeeded H. A. Wilson in the Wheatstone chair of physics in the University of London (King's College), on Wilson's appointment to follow Rutherford at Montreal. Barkla held the Wheatstone professorship until 1913, when he accepted the chair of natural philosophy in the University of Edinburgh. He was able to continue his investigations without serious interruption throughout these changes of domicile because of the simplicity of the apparatus he employed, which did not require a long time for reassembling on starting work in a new laboratory.

In his Bakerian lecture Barkla gave a preliminary report of a new effect attending the passage of a beam of X-rays through certain light elements—an effect to which he gave the name "J-phenomenon", and which formed the subject of most of his investigations after the year 1917. Certain discontinuities in the absorption curves of these elements which occurred for shorter wave-lengths of the primary beam than corresponded to the production of characteristic K and L radiations were at first thought by Barkla to indicate the presence of a new series of characteristic radiations—the J series.

Investigations of the scattered radiation, however, showed that a phenomenon of a different type was involved. It appeared from Barkla's experiments that the primary and the scattered rays might be absorbed to the same extent in passing through a given substance up to a certain critical thickness of that substance, after which the scattered

radiation proved more absorbable in that substance than the primary. However, even after the stage of different absorbability of the primary and the scattered radiation in one material had been attained, the two radiations might show the same penetrating power in a different material. These effects were somewhat erratic in their reproducibility, but their frequent occurrence convinced Barkla that the processes giving rise to the J phenomenon were quite distinct from those responsible either for scattering or for the excitation of fluorescent radiation. Despite the apparently capricious occurrence of the phenomenon, Barkla was confident of its reality. He was convinced that it could not be attributed to any spurious action of the apparatus. In the belief that he was dealing with a fundamentally new effect, he spent much time in attempting to sift the evidence and in trying to elucidate the conditions necessary for obtaining the different effects reproducibly. There is, however, no published record of his having arrived at a complete understanding of the matter.

Barkla was an excellent lecturer, endowed with a powerful sonorous voice which enabled him to address large classes without physical effort. As a teacher he took much trouble to promote the interests of individual students, and was successful in inspiring his abler students with enthusiasm for research work. He was in continual demand in British universities as an examiner in physics and, as a means of supplementing his professorial stipend in the interests of his family, he devoted much time to this work over a period of several years. His varied experience of students in Liverpool, London, and Edinburgh, the wide range of his knowledge of physics, his thoroughness and conscientiousness made him an ideal examiner. He took great pains to arrive at a just decision in the assessment of border-line cases, and his powers of discrimination commanded the respect of his co-examiners.

Barkla became a Fellow of the Physical Society in 1910 and was elected to the Fellowship of the Royal Society in 1912. He served for periods as a member of the council of each of these bodies. The University of Liverpool conferred the honorary degree of LL.D. upon him in 1931.

To many of his contemporaries in Cambridge, Barkla was known chiefly for his remarkable singing voice. It was, indeed, freely predicted that he had a great future before him as a professional singer, and it was sometimes said that he was missing his real vocation in devoting his time to physics. After one year as an undergraduate at Trinity he migrated to King's in order to join King's College choir and have the joyous experience of singing in the magnificent chapel of that college. On occasions when he was due to sing the solo part in an anthem the chapel was invariably crowded. In those days the original Cavendish Laboratory had only been enlarged by the provision of a room for elementary practical classes, and accommodation for research students was limited. Barkla had to carry out his experiments in the cellar of the porter's lodge, which had windows opening into a space underneath a large grating in the footpath of Free School Lane. It was Barkla's habit on arrival in the mornings, after making preliminary adjustments to his apparatus, to burst into song while waiting for things to settle down. In summer time, when the windows of his cellar were open, passers-by, on their way to the various laboratories and museums which were approached through the gateway of the Cavendish, would stop and look around with astonishment at the volume and quality of the vocal efforts which arose apparently from the bowels of the earth!

While a lecturer at the University of Liverpool, Barkla married Mary Esther, elder daughter of the late John T. Cowell, Receiver-General of the Isle of Man, and they had three sons and a daughter. The sudden death of the youngest son, Flight-Lieutenant Michael Barkla, M.B., Ch.B., at Carthage in August 1943 was a great sorrow to the family. He had had a brilliant career both at school and at the university, and seemed destined for a very distinguished future.

In his private life Charles Barkla was a man of firm religious principles. To his wife and family he gave unstinted devotion. To his closest friends he was the embodiment of kindness, loyalty, good humour, and commonsense. It was part of his personal charm that he never lost the capacity to laugh at himself with his friends. He would, for instance, smilingly recount how, a few years ago, his students produced a physical "alphabet" containing the couplet:

"J's a phenomenon known to the Prof.
On Friday its working, on Monday its off."

Recreation for Barkla meant his home or a holiday with his family and the car, preferably in the Highlands. He was physically robust and enjoyed good health until the last year of his life, when, in June 1944, illness compelled him to undergo an operation. He seemed to have recovered his normal health again by the beginning of the present session, but suddenly he collapsed, was ill for a few days, and died on 23 October.

FRANK HORTON.

JOHN RODERICK ENNIS SMITH

JOHN RODERICK ENNIS SMITH was born on 21 April 1918 at London, Ontario. He was educated at the Summerside High School, Prince Edward Island, from 1931 to 1934, and then at Dalhousie University, Nova Scotia, from 1934 to 1938. He took his B.Sc. at that University and was elected to a Rhodes Scholarship, coming into residence at Trinity College, Oxford, in October 1938. He took the D Phil degree in 1941. After the outbreak of war he was engaged on research work in Oxford for the Ministry of Supply, and from 1941 to 1943 held appointments as Demonstrator, first in the University Physical Chemistry Laboratory and later in the Department of Physics. In 1944 he was appointed as Scientific Officer in the Inter-Services Research Bureau. While on holiday in the summer of that same year he met his death by drowning, on 10 August, at one of the notoriously dangerous places on the Cornish coast.

In spite of the early diversion of his scientific work to war problems, he had already shown himself a man of the highest calibre in original work. His published papers are all on physico-chemical subjects, though latterly his interests began to move in the direction of pure physics. His first piece of work was a comparison of the thermal decomposition of acetaldehyde with that of its deuterium analogue. He then made a study of the kinetics of decomposition of gaseous benzaldehyde and of various substituted benzaldehydes. From this he went on to study the inhibition by small quantities of nitric oxide and by propylene of various reactions, including the decompositions in the gaseous phase of aldehydes and ketones. These studies have been of the greatest help in elucidating the problem of the extent to which such reactions depend upon the intervention of free organic radicals, and have helped very considerably in systematizing our knowledge of the kinetics of such changes. He was generally regarded by those who knew him in the Oxford Chemistry School as a man of the finest intelligence. He had a very keen analytical mind, and was not only a skilful experimenter but a good mathematician. For so young a man, working for so short a time amidst so much distraction, he has left a remarkable contribution.

Regret which must be felt, even impersonally, at the loss of such great scientific promise is multiplied immeasurably in those who knew him well. Gentle-natured, unselfish, a good athlete, and with much personal attractiveness, he inspired and gave friendship and affection in abundant measure.

C N H

MARIUS HANS ERIK TSCHERNING

MARIUS H. E. TSCHERNING was born at Oestrup, Denmark, in 1854 and died in Copenhagen on 9 September 1939.

After graduating in medicine at Copenhagen in 1882, where he had studied under E. Hanson-Grut, Tscherning undertook post-graduate study in Paris and became associated with L. Emil Javal at the Ophthalmological Laboratory at the Sorbonne. The laboratory was founded in 1876 for Javal, whose translation into French of Helmholtz's *Physiological Optics* had appeared in 1867. Tscherning became Javal's assistant in 1884 and later, when Javal was incapacitated by blindness, he was appointed Director of the laboratory. He held this position until 1910, when he returned to Copenhagen to succeed Bjerrum as Professor of Ophthalmology, retiring from this post in 1925.

Following the lead of Javal, Tscherning was a great admirer of the work of Thomas Young, and delivered the first Thomas Young oration to the Optical Society in London in 1907, the year in which he was elected to Honorary Fellowship of the Society. His textbook on *Physiological Optics*, first published in Paris in 1898, was translated into

English by Carl Weiland, of Philadelphia, in 1900, at a time when no other English text on the subject was available. Written in lucid style, free from mathematics, the book became the standard students' text for some years. This would appear to account, in large measure, for the considerable reputation that Tscherning achieved in this country.

In his earlier years Tscherning was very active in investigating many problems in physiological optics. He designed an instrument, the ophthalmophakometer, for utilizing the catoptric images of the eye in measuring the curvatures and separations of the ocular surfaces. His work on the aberrations of the eye was rather severely criticized by Gullstrand. He appears to have been the first to recognize the importance of the centre of rotation of the eye in relation to the design of spectacle lenses; and the "Tscherning ellipse", which is the graph obtained by plotting the total power of the lens against the two forms for each power that are free from primary astigmatism, is well known.

It is probably, however, for his work on the mechanism of accommodation that Tscherning is best known. The central position of his theory, first propounded in 1895, was that the general effect of the ciliary muscle is to increase the tension upon the zonule during the act of accommodation for near objects, whereas Helmholtz maintained the opposite view that the tension on the zonule is relaxed. Tscherning also stressed the importance of Young's observation that the curvature of the lens increases over the central portion only. Although he modified his theory in various respects subsequently (1904 and 1909), he did not move from his central position concerning the increased tension on the zonule. Subsequent evidence, including the important contributions of E. F. Fincham associated with the non-uniform thickness of the elastic lens capsule, leaves practically no doubt that the Helmholtz hypothesis is correct.

Tscherning's contributions to the subject of physiological optics, appearing as they did during a period of relative scarcity of original work in the subject, have been of considerable service in stimulating further enquiry.

H. H. EMSLEY.

JOHN KEATS CATTERSON-SMITH

J. K. CATTERSON-SMITH occupied the William Siemens Chair of Electrical Engineering at King's College, London, from 1930 until his death.

After qualifying at Birmingham and obtaining practical experience at the Stafford works of Siemens Bros., he had a distinguished career as a lecturer in Electrical Machinery and Design at the University of Liverpool, at Finsbury Technical College, and at Faraday House.

In 1923 he became Professor of Electrical Engineering at the Indian Institute of Science, Bangalore, where he did much to foster Anglo-Indian goodwill and co-operation. Here, he founded, and for several years edited, the journal *Electrotechnics*.

Among the papers which Catterson-Smith wrote was an early one dealing with the electric propulsion of ships, which earned him a gold medal, and later he read a notable paper on induction motors before the I.E.E. Commutation, and the design of transformers and D.C. machines were the subjects of other early papers.

His talent as a lecturer and his unusual gift for devising mechanical models of electrical phenomena were well demonstrated when, in 1932, he delivered the Faraday Lectures, sponsored by the Institution of Electrical Engineers, under the title "Everyday Uses of Electricity".

Of late years Catterson-Smith interested himself in symmetrical components. He showed some circuits for their separation and measurement at the 1937 exhibition of the Physical Society.

The Professor's health began to deteriorate during the early part of the war, and, after much suffering, he died on 25 January of this year. He will be missed both by his numerous friends and by the generations of students whom he helped so readily.

B. C. L.

REVIEWS OF BOOKS

The Royal Society 1660-1940: A History of its Administration under the Charters, by Sir HENRY LYONS. Pp. x+354. (Cambridge: The University Press, 1944.) 25s.

Whether the clang of the falling weights dropped by Galileo from the leaning tower of Pisa sounded the death-knell of the Aristotelian philosophy is a point that may well be disputed, but it is indubitably certain that the end of the sixteenth and the beginning of the seventeenth century witnessed an outburst of scientific activity which is almost unexampled in history, and with which the names of Galileo in Italy and of Gilbert in England are indissolubly joined. Francis Bacon at least "rang the bell that call'd the wits together", and his Solomon's House, if it did not form a model for the societies which were springing up on every hand, served a useful end in providing material for the discussion of the constitution of a College wherein the New or Experimental Philosophy should be practised.

The precise origins of our own Royal Society are obscure; but it is clear that, amid the drums and trappings of the English Civil War, there existed in London (in 1645 or thereabout) a small group of men disinterested enough to meet weekly at Gresham College or at convenient places nearby to discourse of "Philosophical enquiries, and such as related thereunto; as Physick, Anatomy, Geometry, Astronomy, Navigation, Staticks, Magneticks, Chymicks, Mechanicks and Natural Experiments . . . These meetings in London continued and (after the King's return in 1660) were increased with the accession of divers worthy and Honourable Persons; and were afterwards incorporated by the name of the Royal Society, etc., and so continue to this day". (Dr. John Wallis. Written in January 1696-7.)

July 15th, 1662, is the date of the sealing of the first Charter. This is the date usually recognized as marking the Society's foundation.

The young Society had the chequered career that falls to the lot of many young Societies. Subscriptions fell into arrears; officials resigned—sometimes in what appears to have been a temper; the Council embarked upon unfortunate publishing ventures, and found itself with many unsold copies of Willoughby's *De Historia Piscium* left upon its hands. Thereafter, the Council minutes make not unamusing reading today, though the humour would not be so evident to contemporary members of the Council. The unlucky paid officers of the Society were offered their arrears of salary in "Books of Fishes", and Halley, set to the task of measuring the length of a degree of longitude, was offered a payment of fifty pounds or fifty Books of Fishes. The Society was indeed fortunate, during this difficult period, in possessing a clerk of the calibre of Halley, and doubly fortunate in that it was in the main due to Halley that the *Principia* was written and published. If, in the long and distinguished history of the Society, there is one day, above all others, deserving of commemoration, it is that August day, in the Cambridge of 1684, when Halley asked of Newton what would be the path of a planet under a force varying inversely as the square of the distance, and received the reply, "An ellipse". That question and reply largely determined the course of physical science during the two centuries that followed. For it was Halley who urged Newton to the continuation of his work; who overcame Newton's constitutional aversion to publication; who saw the work through the press, and who published the book at his own charges, for the printer could hardly be expected to accept payment in Books of Fishes! Halley's claims to the gratitude of the scientific world are many; but his share in the publication of the *Principia* is by far the greatest.

The temptation to enlarge upon the many dramatic incidents in the Society's history is difficult to resist, and it is pleasant to note that Sir Henry has not been unmindful of such incidents. In particular he tells in some detail the story of the lightning-conductor controversy. In 1772, the Council of the Society appointed a committee, of which Benjamin Franklin was a member, to report upon the best form and disposition to be

given to lightning conductors ; the committee reported in favour of pointed conductors—the form introduced by Franklin—one member (B. Wilson) dissenting and expressing an opinion in favour of blunt conductors. In 1777 the opinion of the reconstituted committee was again sought, with like results. But the war of the American Revolution had by this time broken out, and the question became a political one—advocates of blunt conductors were regarded as patriotic Britons, of pointed conductors as revolutionaries !

It is credibly reported that King George III interviewed the President of the Royal Society (Sir John Pringle), and requested the Society to reconsider its decision, and to report in favour of blunt conductors, whereupon the President remarked that laws of Nature were not alterable at the Royal pleasure. At the Anniversary Meeting of 1778 Sir John Pringle did not offer himself for re-election, *post hoc* and possibly *propter hoc*.

There are few figures in the annals of eighteenth-century science more intriguing than that of John Hill (1716–1775), and it is surprising, and indeed rather disappointing, to find him curtly dismissed by Sir Henry as “one John Hill, a quack doctor”.

Hill was versatile and a man of immense industry. Botanist, actor, dramatist, translator and doctor (his degree was M.D. of St. Andrews), he may have been most things by turns, but he was faithful to his botanical studies. He was heartily disliked by many of his contemporaries, and there is a temptation to regard as a true picture Garrick's cutting epigram :

“For farces and physic his equal there scarce is,
His farces are physic ; his physic a farce is.”

Or the less well known

“Thou essence of dock, valerian, and sage,
At once the disgrace and the pest of this age,
The worst we can wish thee for all thy damn'd crimes,
Is to take thy own physic, and read thy own rhymes.”

Which was at once capped by

“Their wish must be in form reversed to suit the doctor's crimes ;
For if he takes his physic first, he'll never read his rhymes.”

All this goes to support the contention that Hill was vain, provoking and quarrelsome ; but when his contributions to botany are examined, they are seen to be real and substantial. He knew his plants intimately and widely, he did useful work in plant physiology and in taxonomy ; he wrote an account of the vegetable kingdom in *twenty-six volumes folio* ; a more modest effort, his *British Herbal* in one volume folio, is a mine of curious information, and it would seem hardly just to dismiss as a quack doctor a man who has been selected, with Ray, Grew, Hales, Hooker, Henslow and Marshall Ward, as one of the Makers of British Botany. His quarrel with the Royal Society was probably due to his failure to obtain the necessary signatures to his certificate of candidature, but it resulted in the publication of his *Review of the Works of the Royal Society of London* (1751)—a work which pillories the more feeble papers appearing in the *Philosophical Transactions* to that date, and which was probably no inconsiderable factor in producing a higher standard of publication in the *Transactions*.

While dealing with the matter of minor omissions, it may not be amiss to draw attention to the career of Walter White, assistant secretary to the Society from 1861 to 1883, who receives no more than bare mention of his name. White was an earnest (and humourless) Victorian, whose character had been developed in the tradition of Samuel Smiles, and his *Journals*, published in 1898, after his death, furnish a most interesting (and, on occasion, surprising) series of thumb-nail sketches of celebrities of the nineteenth century.

But these are side-issues. An authoritative history of the Society was long overdue—Weld's history, the immediate predecessor of the present volume, was published in 1848, and brought the story down to 1830. Sir Henry's book is primarily an “account of the way in which the Society carried on its work” ; it is based on a close and critical study of the Council minutes, and its author regards it as a preliminary study to a complete history of the Society. This may indeed be the case ; but it would seem to be somewhat in the nature of a meiosis to rate as a preliminary study a work which gives a clear and comprehensive picture of the founding of the Society ; of its early struggle for existence ;

of the period of prosperity under Newton's long reign ; of the time when the scientific work of the Society was in danger of being swamped by the number of wealthy amateurs and dilettanti recruited to the ranks of the Society ; of the dictatorship of Sir Joseph Banks ; of one period which Sir Henry terms that of the scientific revolt (1820-1860) ; and that of the later years when, as De Morgan said, "a Fellow of the Society should be a Fellow Really Scientific". It is a great story, and a story which loses nothing in Sir Henry's sympathetic hands. There is much more in the book than a cold record of administrative detail, as we have seen, dramatic incidents in the Society's history are not passed over, and, period by period, there are given admirable biographical sketches of Presidents and Senior Officers of the Society.

It is a sad thought that Sir Henry, who had passed the book for the press, did not live to see its publication ; but he has, all unknowing, reared an enduring monument, and his labours have earned for him the affectionate thanks of all students of the history of science.

A. F.

Handbook of Industrial Radiology, by Members of the Industrial Radiology Group of the Institute of Physics. Edited by J. A. CROWTHER. Pp. viii + 203. (London : Edward Arnold and Co., 1944.) 21s. net.

This handbook, which "had its origin in a series of lectures given before the Industrial Radiological Group of the Institute of Physics", is a timely production. Although the last twenty years have seen no revolutionary changes in x-ray apparatus and technique, progress has been steady and solid, and the often meretricious exterior of the modern set conceals a triumph of material and design. Part of the story of its development may be guessed as this book is read, but the more particular purpose of the authors has been to state the principles and describe the technique of modern industrial radiology. The result is a volume which should find its place on every radiographer's bookshelf.

The book comprises nine chapters, uneven in quality, each dealing with some particular aspect of the subject. Editing has reduced overlap to a negligible amount.

Three chapters are particularly good. In that on the physical principles underlying radiography the author has made a good choice of material and, in the space available, has succeeded, without sense of overcrowding, in including the essentials in a most interesting account. The chapter on *The response of photographic materials to x-rays* is well above the usual standard of writing upon that subject. The factors involved in the production of a good negative are unusually well and logically stated. Gamma radiography is the subject of another good chapter. The author has taken full advantage of the fascinating material he has had to present. Radium (or radon) may be expected to find an increasing use in the future, particularly because of the possibility of using it in situations quite inaccessible to an x-ray tube.

An adequate description of modern x-ray equipment is given in a chapter headed *Requirements in design of industrial x-ray equipment*. Improvements could have been effected in the descriptions of high-tension circuits. The author draws attention to the short lives which tubes have when operated at constant high potential ; there seems to be a simpler reason for this than he suggests. The features (good and bad) of modern cathode design would have repaid treatment. The typical industrial set is well described and a short description is given of the megavolt sets now coming into use.

A considerable fraction of the chapter on quantitative measurements is devoted to dosage. Few subjects provide so much evidence of confused thinking. From this the author is more than usually free, though an unhappy phrase in one place might reinstate the confusion in the reader's mind.

The chapter on the radiography of heavy metals prescribes a technique which the beginner might reasonably follow, though, as he gains experience, he will in some respects depart from it with advantage. The author's paragraph on the reporting of radiographs is interesting ; one would wish it longer.

Even with the standard of reproduction of this book, radiographic flaws should be indicated by arrows. It is unfortunate that the author of the chapter on the radiography of light metals should point out how the definition of a flaw is improved by filtration when in the illustration one can only guess which is the flaw and which are blemishes

in reproduction. It is also disturbing to find that a great improvement has resulted from an increase in voltage when the author points out that the opposite has occurred. The particular features of light metals from the radiographic point of view are adequately treated.

A chapter on uncommon applications of industrial radiography contains a good deal of material of popular interest. A chapter on x-ray protection concludes the book.

The production is good, especially judged by war-time standards. Very few misprints occur. The most serious are due to alignment of quotients in exponents, causing transfer of terms from numerator to denominator. Three such cases were noticed. A. G. W.

Sound Insulation and Acoustics, by the Acoustics Committee of the Building Research Board of the Department of Scientific and Industrial Research. Pp. 80. (London: H.M. Stationery Office, 1944.) 1s

This is one of a series of pamphlets dealing with various aspects of post-war building. In a methodical way it describes current practice in the acoustical treatment of buildings, both from the point of view of preventing unwanted sound from entering a building, or passing from one room to another, and from that of obtaining desirable conditions for speaking and making music within a room. Valuable features of the pamphlet are the tables giving concise data on the noise reduction to be expected with various types of building construction and of up-to-date values of absorption coefficients of materials.

As most of the research work on sound during the war has been along lines unconnected with buildings, the reader must not expect to find much that is new. One novel suggestion is to screen the lower floors of buildings in town from street noise by planting shrubs and trees of thick foliage along the verge of the road, though, as the Committee point out, experiments are needed to show how effective such a screen would be.

The Committee is sometimes a little too optimistic as to the economic aspects of sound insulation. They consider that, for example, the constructions they advocate for insulating blocks of flats will be inexpensive, but this term must be taken in relation to pre-war conceptions. Certainly the specially isolated rooms for music practice in the home, which they propose, can never be inexpensive in relation to the urgency of their incorporation in houses and flats. A generation that has become indifferent to radios blaring into courtyards and gardens in the summer is not going to cavil at little Mary practising her scales in the adjacent dwelling.

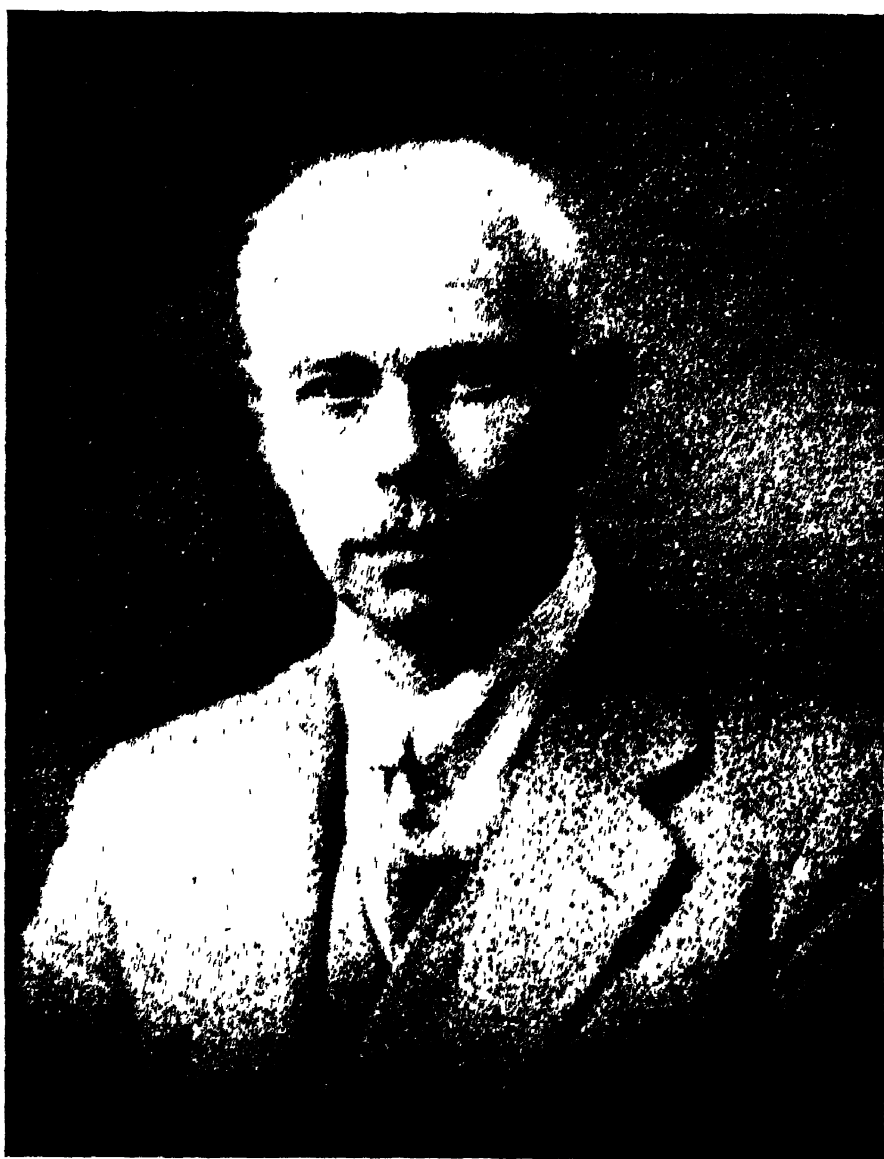
This pamphlet can be recommended to all who want what is in fact an up-to-date and concise text of all the essentials of "Acoustics of Buildings". At a shilling, it must make all private publishing firms who have a book on the subject on their lists gnash their teeth! E. G. R.

RECENT REPORTS AND CATALOGUES

Supplementary List of Publications of the National Bureau of Standards, January 1 1942 to June 30 1944 (with Subject and Author Indexes), compiled by J. L. MATHUSA. (Supplement to Circular C24, November 1944.) Pp. 84. U.S. DEPARTMENT OF COMMERCE, Washington, D.C. 20 cents.

Abridged Scientific Publications from the Kodak Research Laboratories, Volume 25, 1943. Pp. 443+xii. EASTMAN KODAK COMPANY, Rochester, N.Y.

A Report to the Government of India on Scientific Research in India, by Professor A. V. HILL, M.P., Sc.D, F.R.S. Pp. 55. Reprinted in the United Kingdom for private circulation. THE ROYAL SOCIETY, Burlington House, London W. 1.



F W ASTON, F.R.S
Twenty-first Duddell Medallist

(See back of plate.)

F. W. ASTON, F.R.S., TWENTY-FIRST DUDELL MEDALLIST

The twenty-first Duddell Medal was presented to Dr. Aston on 23 May 1945 by the President, Prof E. N. da C. Andrade, F.R.S.

The medallist, who is 67, was educated at Malvern College and Mason College, and at the Universities of Birmingham and Cambridge.

He is known to all physicists for his invention, development and application of the mass spectrograph, and it is for this work that the medal was awarded. His mass spectrograph, first designed and constructed in 1919, made use of a new method of electromagnetic focusing, which, he revealed on the occasion of the presentation, was worked out numerically for a particular case, a method which he preferred to that of calculating algebraically a more general case. With the instrument and its successors he established first that most elements consist of mixtures of isotopes, later that the masses of these isotopes are very close to integers if the mass of oxygen is taken as 16, and then that the differences from whole numbers varied in a systematic, and not in an entirely random way. For this last work, using a much improved spectrograph, he has measured masses with a precision of 1 part in 20,000, and thus determined the binding energy of light elements with great accuracy. By measuring also the relative abundances of the different isotopes, Aston is enabled to compute "chemical atomic weights" with precision.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 57, PART 4

1 July 1945

No. 322

DEFINITIVE EQUATIONS FOR THE FLUID RESISTANCE OF SPHERES

By C. N. DAVIES,

Industrial Health Research Board, Medical Research Council, London

MS. received 22 January 1945

ABSTRACT. For calculation of terminal velocities it is convenient to express the Reynolds' number, Re , of a moving sphere as a function of the dimensionless group ψRe^2 , where ψ is the drag coefficient. The following equations have been fitted by the method of least squares to critically selected data from a number of experimenters

$$1 \quad Re = \frac{\psi Re^2}{24} - 0.00023363(\psi Re^2)^2 + 0.0000020154(\psi Re^2)^3 - 0.0000000069105(\psi Re^2)^4$$

for $Re < 4$ or $\psi Re^2 < 140$. This tends to Stokes' law for low values of Re . It is specially suited to calculation of the sedimentation of air-borne particles. The upper limit corresponds to a sphere weighing $1.5 \mu\text{g}$. falling in the normal atmosphere, that is, one having a diameter of 142μ for unit density.

$$2. \log Re = -1.29536 + 0.986 (\log \psi Re^2) - 0.046677 (\log \psi Re^2)^2 + 0.0011235 (\log \psi Re^2)^3$$

for $3 < Re < 10,000$ or $100 < \psi Re^2 < 4.5 \cdot 10^7$

Correction for slip in gases should be applied to Stokes' law by the following expression, based on the best results available.

$$1 + \frac{l}{a} [1.257 + 0.400 \exp(-1.10a/l)],$$

where the mean free path l is given by $\eta/0.499\sigma\bar{c}$

This conveniently transforms to the following for the sedimentation of particles in air at pressure p cm. mercury.

$$1 + \frac{1}{pa} [6.32 \cdot 10^{-4} + 2.01 \cdot 10^{-4} \exp(-2190ap)].$$

§ 1 INTRODUCTION

A COMMON requirement in the interpretation of experiments is a knowledge of the fluid resistance offered to the motion of spheres; work on viscosity, or involving the sedimentation of dusts or liquid suspensions, can be cited, and there are many other applications. Experiments on this subject have been carried out since the time of Newton and there exists a large bulk of material upon which various empirical equations have been based. None of these, however, is derived quantitatively from selected data, and most suffer from the disadvantage that explicit solution for the terminal velocity of a falling sphere is impossible.

The present attempt provides equations which do not exhibit this defect. It is not possible to deduce a single formula, covering all sizes and speeds, without overburdening it with constants; the wide range which must be covered, for

practical reasons, has, therefore, been split into two. The equation covering the lower section tends to Stokes' law, in the limit, and at the same time is accurate for spheres of any size, large compared with the mean free path, and weighing less than $1.5 \mu\text{g.}$ (diameter 142μ for unit density) falling freely in normal air; hence it is useful for most problems concerned with air-borne particles and also provides the best available assessment of viscosity experiments. The other equation continues, using logarithms on account of the large range, up to a Reynolds' number equal to 10,000, above which value the interest becomes purely aerodynamical. This region covers all spheres falling through normal air which have masses between $1.1 \mu\text{g.}$ and 0.48 g.

Finally, the effect of slip for motion through gases in the Stokes' law region is discussed in the light of the best available experimental data

§ 2. THEORETICAL BASIS

A sphere of diameter d moving steadily through a medium of density σ and viscosity η , with velocity v , experiences a resistance force W . Hence, if no other variables are concerned, the problem is summarized by the equation

$$f(d, v, \sigma, \eta, W) = 0. \quad \dots\dots(1)$$

By Buckingham's theorem (1914) this can be reduced to

$$f\left(\frac{vd}{\nu}, \frac{8W}{\pi d^2 \sigma v^2}\right) = 0, \quad \dots\dots(2)$$

where $\nu = \eta/\sigma$ is the kinematic viscosity of the fluid and the factor $8/\pi$ is inserted for conventional reasons. vd/ν is the Reynolds' number, Re , and $8W/\pi d^2 \sigma v^2$ is the drag coefficient, ψ . The latter can be regarded as a variable coefficient which must multiply the product of cross-section ($\pi d^2/4$) and dynamic pressure ($\frac{1}{2}\sigma v^2$) in order to yield the resistance: hence the numerical factor.

Convention has established the practice of exhibiting experimental results as graphs of Re against ψ . This is more convenient than plotting the actual measurements, such as velocity and diameter, since, by equation (2), a single curve must be obtained for all possible variations of fluid and sphere, providing the conditions of equation (1) are not transgressed. Also, because Re and ψ are dimensionless, any consistent system of units may be used in the measured quantities without altering their magnitudes. However, if it is desired to use such a curve to derive velocities, indirect procedure is necessary since v occurs in both variables.

This obstacle is avoided by arranging equation (2) thus:

$$f(Re, \psi Re^2) = 0, \quad \dots\dots(3)$$

since $\psi Re^2 = 8W/\pi \sigma v^2$, which contains only properties of the fluid. In the case of a sphere rising or falling freely at its terminal velocity we have

$$W = (m - m')g,$$

m being its mass and m' the mass of fluid displaced. Hence if Re is expressed by an equation in ψRe^2 , the independent variable involves only the mass of the sphere and the nature of the fluid, so that direct solution for v is possible.

Stokes' law of resistance can be reduced to

$$Re = \psi Re^2/24, \quad \dots\dots(4)$$

This is well established for very small values of Re , but it forecasts too low a resistance as Re increases. Hydrodynamical equations due to Oseen and Goldstein are valid a little further than that of Stokes, but then yield resistances which are too great. The precise determination of a point at which experiments show these equations to be in error by a definite amount is a matter of difficulty. The equation of Goldstein is cumbersome and apparently little better than that of Oseen, while the latter's suffers from the disadvantage that it leads to a quadratic in Re when ψRe^2 is made the independent variable. It is therefore inconvenient to fit an empirical polynomial in ψRe^2 to the experimental results which would tend, as Re diminished, to Oseen's expression. For this reason the procedure adopted has been to fit selected observations, by the method of least squares, with a polynomial in ψRe^2 which passes through the origin of a plot of Re against ψRe^2 and which has, at the origin, the slope given by equation (4). In other words it tends to Stokes' law as Re tends to zero. In order to give a satisfactory fit for Re less than 4, a quartic in ψRe^2 has been found necessary; this leaves three coefficients to be determined by least squares, since there is no absolute term and the coefficient of the first-power term is $1/24$.

For $Re > 4$ a cubic, using the logarithms of the same variables, has been deduced. This contains four numerical constants. It is thus not possible to make it tend to Stokes' law, in the limit, and it must not be used for lower Reynolds' numbers. The upper limit of validity is $Re = 10^4$, which is a little beyond the point at which the drag coefficient passes through a minimum. Experimental data at higher Reynolds' numbers are not discussed.

§3. OBSERVATIONS WITH REYNOLDS' NUMBERS LESS THAN 4

The latest determinations are due to Möller (1938) and cover the range $0.0507 < Re < 1.603$. His results are corrected for wall effect by Faxén's method, and he was at pains to keep the correction small; it ranges from 0.02% to 0.81%, and the ratio of sphere diameter to the diameter of the vessel is always below 0.02. Schmiedel (1928) covers the range $0.053 < Re < 1.493$ and uses the Faxén correction, which varies between 1.1% and 7.7%, the ratio of sphere to tube diameter being between 0.018 and 0.074. Liebster (1927) has results for $0.1355 < Re < 1.927$, also evaluated by Faxén's formula. The maximum wall correction is 2.68% and the least 0.21%, with diameter ratios up to 0.0525. Seven results for $2.32 < Re < 4.16$ have been used from a series of experiments extending to higher Reynolds' numbers. The diameter ratios were all below 0.0353, and no correction has been applied for wall effect.

The Faxén correction is based upon Oseen's equation, which is increasingly erroneous as Re exceeds unity. However, with increasing Re , the correction diminishes and is probably quite negligible in this instance. There is no experimental work on wall effect from $Re \approx 2$ until we come to determinations by Lunnon in the region of $Re \approx 10^4$. For a diameter ratio 0.02 at $Re = 0.4$ the Faxén correction to the drag coefficient is $1\frac{1}{2}\%$; at $Re = 10^4$ an equation given by Lunnon yields a 1% correction for diameter ratio 0.022. It appears, therefore, that over a very wide region the wall effect may be small and approximately constant.

Use of the Ladenburg formula for wall effect gives a correction which is independent of Reynolds' number. Thus, as the Reynolds' number increases,

too great an allowance is made for the increased resistance due to wall effect; experimental data are thereby shifted to show specious agreement with Stokes' law, though coincidence at lower values, which would be genuine, may be impaired. The data of Arnold (1911) show this defect. Many of his determinations involved a large wall correction. Lemin (1931) covered the range $0.00028 < Re < 0.0053$; he showed, in this region of very low Reynolds' numbers, that the Stokes-Ladenburg formula was valid if the diameter ratio was below 0.06. We are not concerned with Lemin's figures in the present discussion, since their function is the confirmation of Stokes' law, but Arnold has results for $0.002 < Re < 1.9$. These have been recalculated by Schiller (1932) with the Faxén correction, and we have adopted for use in this paper those having a diameter ratio below 0.06, following Lemin, and within the range $0.05 < Re < 0.19$.

There remains for consideration the work of Allen (1900). Timing the ascent of air bubbles in aniline and water, he covered a range from $Re = 0.009$ up to 25. The results are scattered, relative to modern measurements, and, while of historical importance, do not merit inclusion in our analysis. Some measurements were also done with spheres of paraffin wax rising through aniline for $0.554 < Re < 20.2$. Drag coefficients calculated from these are consistently high relative to those from the workers previously mentioned, so these figures are likewise ignored.

We are left with determinations due to Möller, Schmiedel, Liebster and Arnold. Computed values of ψRe^3 written in order of Reynolds' number are given in table 1 and number 68 in all. The lower limit of Reynolds' number was selected as 0.05, since below this the failure of Stokes' law is negligible. Möller's figures deviate progressively from those of Schmiedel for $Re > 0.2$ in the direction of low resistance. Consistent error of 1% in viscosity determinations, a possibility which Moller states may exist in his own work, could not account for this. Liebster's points, on the whole, lie nearer to Möller's than to those of Schmiedel. The latter has been criticized because of the magnitude of the wall correction applied in some cases (8%), but these individual experiments do not fall out of the general sequence which is established by many having only a 2% correction. Liebster's correction exceeded 2% in only one instance. Moller was careful to keep the correction small because he wished to demonstrate whether experiments conformed more closely to Oseen's law than to Stokes'. His conclusion would have been unconvincing if the correction had been larger than the small difference in question.

There seems to be no reason for weighting the determinations of these four workers unequally, accordingly their observations have been fitted by the method of least squares with the following equation:

$$Re = \psi Re^3 / 24 - 0.00023363 (\psi Re^2)^2 + 0.0000020154 (\psi Re^2)^3 - 0.000000069105 (\psi Re^2)^4. \dots (5)$$

This, therefore, represents the best available means of calculating terminal velocities for $Re < 4$ or $\psi Re^3 < 140$.

The fit of this equation and the scatter of the observations can be judged by the residuals, shown in table 1, and by the quantities in table 2. Owing to the wide range of Re values (eighty-fold) the residuals would be greater for the higher

Table 1

Author	Re	ψRe^a	Calculated Re	Residuals, v Exp. — Calc.	v^2
1 M	0.0507	1.235	0.0531	—4 10^{-4}	16 $\cdot 10^{-8}$
2 S	0.0531	1.296	0.0536	—5	25
3 A	0.0538	1.288	0.0533	5	25
4 A	0.0624	1.472	0.0608	16	256
5 A	0.0771	1.873	0.0772	—1	1
6 M	0.0865	2.107	0.0868	—3	9
7 M	0.0882	2.130	0.0877	5	25
8 M	0.1000	2.431	0.0999	1	1
9 M	0.1008	2.453	0.1008	0	0
10 S	0.1008	2.503	0.1029	—21	441
11 S	0.1059	2.610	0.1072	—13	169
12 A	0.1090	2.613	0.1073	17	289
13 L	0.1355	3.292	0.1347	8	64
14 A	0.1450	3.575	0.1461	—11	121
15 A	0.1510	3.625	0.1481	29	841
16 S	0.1515	3.801	0.1551	—36	1296
17 S	0.1537	3.855	0.1573	—36	1296
18 A	0.1570	3.870	0.1579	—9	81
19 L	0.1600	4.024	0.1640	—40	1600
20 S	0.1636	4.084	0.1664	—28	784
21 M	0.1703	4.213	0.1715	—12	144
22 M	0.1730	4.256	0.1733	—3	9
23 A	0.1900	4.658	0.1892	8	64
24 M	0.2016	5.004	0.2029	—13	169
25 M	0.2038	5.054	0.2049	—11	121
26 S	0.2437	6.242	0.2515	—78	6084
27 M	0.2687	6.723	0.2702	—15	225
28 M	0.2751	6.776	0.2722	29	841
29 L	0.3160	7.719	0.3086	—26	676
30 S	0.3287	8.470	0.3373	—86	7396
31 M	0.3847	9.779	0.3869	—22	484
32 M	0.3914	9.848	0.3895	19	361
33 S	0.4755	12.490	0.4877	—122	14884
34 S	0.4924	12.933	0.5040	—116	13456
35 S	0.4953	13.043	0.5080	—127	16129
36 L	0.5110	12.663	0.4942	168	28224
37 S	0.5157	13.702	0.5320	—163	26569
38 M	0.5547	14.328	0.5547	0	0
39 M	0.5592	14.416	0.5579	13	169
40 L	0.5660	15.443	0.5948	—288	82944
41 L	0.5680	14.646	0.5662	18	324
42 M	0.6464	16.910	0.6470	—6	36
43 M	0.6491	17.013	0.6506	—15	225
44 S	0.6693	18.377	0.6985	—292	85264

Table 1—*continued*

Author	Re	ψRe^2	Calculated Re	Residuals, v Exp — Calc.	v^2
45 L	0 6810	17 810	0 6787	23	529
46 S	0 7277	20 275	0 7644	—367	134689
47 S	0 7489	19·570	0 7400	89	7921
48 S	0 7640	19 620	0 7418	222	49284
49 M	0 8389	21 996	0 8233	156	24336
50 S	0·9241	26 08	0 9615	—374	139876
51 S	1 005	27 79	1 017	—120	14400
52 S	1 042	28 95	1 055	—150	22500
53 S	1 087	31 13	1·125	—380	144400
54 M	1·169	31 29	1·130	390	152100
55 L	1 242	35 64	1 268	—260	67600
56 L	1 492	40 67	1 425	670	448900
57 S	1·493	42·93	1·494	—10	100
58 M	1·501	41 30	1·444	570	324900
59 L	1·530	42·18	1 471	590	348100
60 M	1 603	45·04	1·551	520	270400
61 L	1·927	56·33	1·896	310	96100
62 L	2·32	70·61	2·315	0	0
63 L	2·53	79·05	2·560	—300	90000
64 L	3·005	96·35	3·053	—480	230400
65 L	3·94	132·54	3·978	—400	160000
66 L	3·94	129·75	3·918	200	40000
67 L	4·07	138·28	4·097	—300	90000
68 L	4·16	139·35	4·118	400	160000

Table 2. Analysis of residuals. Equation (5). $0\ 0507 < Re < 4\cdot16$

Author	Number of observa- tions	Σv	Σv^2	Standard error σ	\bar{Re}	σ, \bar{Re}
Möller	21	0·1599	0·00774571	0 0192	0 478	0·040
Schmiedel	21	—0 2213	0·00686963	0 0194	0·559	0·035
Arnold	8	0·0054	0·00001678	0 00145	0 118	0 012
Liebster	18	0·0293	0·01845461	0·031	1·836	0·017
All	68	—0 0267	0·03308673	0·022	0·822	0·027

Reynolds' numbers, although the percentage error in the experiments remained nearly constant. For this reason the standard error for each experimenter is given both alone and also after division by the mean of his Reynolds' numbers.

The curve defined by equation (5) can be seen from the sums of the residuals

to lie between the data of Moller and Schmiedel, but a little nearer to those of the former. Liebster's results govern its course at the upper end of the range.

In using this equation it may be noted that the fourth-power term is negligible, as far as the fourth decimal place, for Reynolds' numbers below 0.35. Below 0.11 the cube can be neglected. For the third decimal place the figures are, respectively, 0.6 and 0.25.

In table 3 the differences between Reynolds' numbers from the equation and those calculated by Stokes' law are given. The latter predicts numbers which are too great.

Table 3. The difference between Stokes' law and equation (5)

<i>Re</i>	0.82	0.38	0.15	0.074	0.037
% difference in <i>Re</i> (Stokes' law is high)	10%	5%	2%	1%	0.5%

For use in viscosity determinations it would be best on physical grounds, as well as convenient, to work at low Reynolds' numbers so that the cubic and fourth-power terms are negligible. A cubic equation in the kinematic viscosity then results which leads to a solution in which this quantity is given explicitly in terms of the experimentally determined variables.

§ 4. REYNOLDS' NUMBERS BETWEEN 4 AND 10⁴

Experiments within this range have been carried out by Liebster (1927), Davies (1939), Wieselsberger (1922, 1923), Lunnon (1928) and Allen (1900). Results due to the last-named are not included. Those obtained with amber spheres, having Reynolds' numbers below 204, yield drag coefficients which are above the general trend of other workers' values and some with steel spheres, with $2304 < Re < 8247$, as computed by Castleman (1926), Wieselsberger (1922, 1923) and the present writer, come out too low. Liebster (1927), however, has obtained different values in some way and shows them well in line with the determinations of Wieselsberger and Lunnon. Schiller (1932) reproduces his figures.

Liebster's own figures for Reynolds' numbers exceeding 1000 have been abandoned since the scatter is great and the trend doubtful, for reasons explained by the author. Seven observations for Reynolds' numbers down to 2.32 are included to provide an overlap with the previous equation. From 1.927 up to 32.6 only Liebster's experiments have been used.

The author's own work was carried out in air with spheres of paraffin wax and small liquid drops; these fall like rigid spheres if not too large. A paper will be published in due course.

The experiments of Wieselsberger consisted of direct drag measurements on spheres suspended in a wind tunnel. All falling within our upper limit have been considered. The same applies to Lunnon's accurate data on spheres falling in water.

Values of $\log Re$ and $\log \psi Re^2$ computed from these various sources are given in table 4. No wall corrections have been attempted since in all cases they would

Table 4

Author	$\log Re$	$\log \psi Re^2$	Calculated $\log Re$	Residuals, v Exp.—Calc.	v^2
1 L	0.366	1.849	0.375	— 9	81
2 L	0.403	1.898	0.416	—13	169
3 L	0.478	1.984	0.486	— 8	64
4 L	0.596	2.122	0.597	— 1	1
5 L	0.596	2.113	0.590	6	36
6 L	0.610	2.141	0.613	—3	9
7 L	0.619	2.144	0.615	4	16
8 L	0.669	2.195	0.657	12	144
9 L	0.849	2.452	0.858	—11	121
10 L	0.886	2.485	0.884	2	4
11 L	1.004	2.644	1.006	— 2	4
12 L	1.083	2.723	1.066	17	289
13 L	1.106	2.788	1.115	— 9	81
14 L	1.165	2.852	1.163	2	4
15 L	1.283	3.027	1.293	—10	100
16 L	1.383	3.150	1.382	1	1
17 L	1.513	3.325	1.508	5	25
18 D	1.591	3.404	1.564	27	729
19 L	1.744	3.659	1.743	1	1
20 D	1.887	3.869	1.886	1	1
21 L	1.918	3.913	1.915	3	9
22 D	2.097	4.196	2.103	—6	36
23 L	2.159	4.274	2.154	5	25
24 D	2.200	4.350	2.203	—3	9
25 D	2.209	4.337	2.195	14	196
26 D	2.213	4.369	2.215	—2	4
27 D	2.263	4.442	2.262	1	1
28 D	2.268	4.449	2.266	2	4
29 L	2.398	4.658	2.398	0	0
30 D	2.515	4.844	2.513	2	4
31 L	2.528	4.850	2.517	11	121
32 L	2.707	5.164	2.706	1	1
33 L	2.786	5.343	2.812	—26	676
34 L	2.843	5.407	2.849	— 6	36
35 W	2.898	5.466	2.883	15	225
36 L	2.926	5.535	2.923	3	9
37 D	2.935	5.558	2.936	— 1	1
38 W	3.061	5.813	3.080	—19	361
39 D	3.086	5.835	3.092	— 6	36
40 Lun.	3.149	5.947	3.154	— 5	25
41 W	3.170	5.973	3.168	2	4
42 W	3.267	6.167	3.274	— 7	49
43 W	3.324	6.276	3.332	— 8	64
44 Lun.	3.364	6.335	3.363	1	1

Table 4—continued

Author	$\log Re$	$\log \psi Re^2$	Calculated $\log Re$	Residuals, v Exp.—Calc.	v^2
45 W	3.373	6.386	3.390	-17	289
46 W	3.500	6.583	3.493	7	49
47 W	3.544	6.690	3.548	-4	16
48 W	3.597	6.780	3.594	3	9
49 Lun.	3.645	6.873	3.641	4	16
50 W	3.721	7.024	3.717	4	16
51 Lun.	3.829	7.241	3.823	6	36
52 W	3.841	7.269	3.837	4	16
53 W	3.924	7.446	3.922	2	4
54 W	4.009	7.627	4.008	1	1
55 Lun.	4.037	7.671	4.029	8	64

probably have been negligible. There are thus 55 observations for $0.366 < \log Re < 4.037$. These have been fitted by the following polynomial:

$$\log Re = -1.29536 + 0.986 (\log \psi Re^2) - 0.046677 (\log \psi Re^2)^2 + 0.0011235 (\log \psi Re^2)^3 \dots (6)$$

The residuals are shown in table 4 and the standard errors, etc., in table 5. They relate, in this case, to the logarithms of the Reynolds' numbers.

Table 5. Analysis of residuals (logs). Equation (6). $0.366 < \log Re < 4.037$

Author	Number of observations	Σv	Σv^2	Standard error
Liebster	26	-0.025	0.002027	0.0088
Davies	11	0.010	0.001021	0.0096
Wieselsberger	13	-0.017	0.001103	0.0091
Lunnon	5	0.014	0.000142	0.0053
All	55	-0.018	0.004293	0.0088

Table 6. Overlap of equations (5) and (6). Liebster's data

Experimental Re	Calculated Re		Eq. (6)—Eq. (5)
	(5)	(6)	
2.32	2.315	2.371	0.056
2.53	2.560	2.606	0.046
3.005	3.053	3.062	0.009
3.94	3.978	3.954	-0.024
3.94	3.918	3.890	-0.028
4.07	4.097	4.102	0.005
4.16	4.118	4.121	0.003

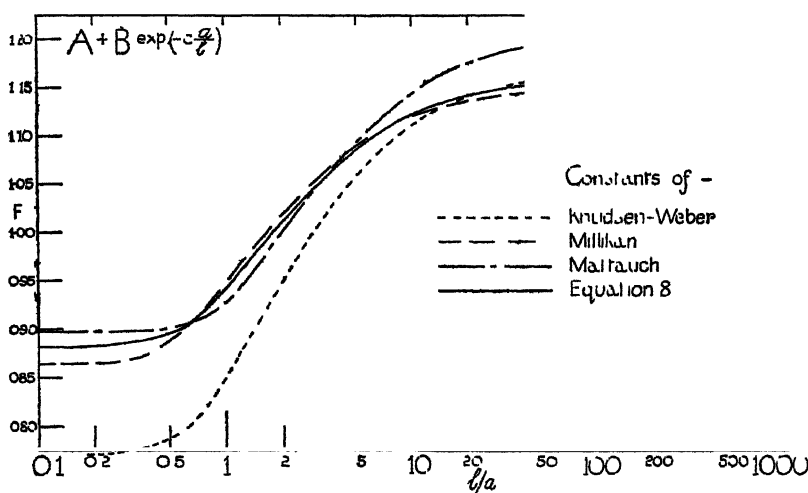
The agreement between the two formulae in the region of overlap is demonstrated in table 6. It is clear that equation (6) can be used down to $Re = 3$ to give values of Re within 1% of those from equation (5).

§5 MOTION IN GASES THE SLIP EQUATION

In the case of a sphere falling through a gas, the density of which is small in comparison, Stokes' law predicts that the terminal velocity will be independent of pressure as a result of the constancy of viscosity, which is accounted for by kinetic theory. When the gas is at a sufficiently low pressure it can no longer be regarded as a continuous medium with respect to the sphere, which then falls more rapidly than hydrodynamic theory would suggest. This is explained as due to slip of the gas at the surface, since the mean free path of its molecules has become comparable with the size of the sphere, and one of Stokes' basic assumptions is violated. Having considered the departure from his law with increasing Reynolds' number it is now proposed to examine results in the other direction.

Knudsen and Weber (1911) carried out experiments with glass spheres 0.389 cm. radius, fixed to a suspended beam, by observing the damping of torsional oscillations in air. They worked from atmospheric pressure down to 0.14 dynes/cm² and deduced an equation for slip correction the form of which has been retained by all subsequent experimenters:

$$F = 1 + \frac{l}{a} [A + B \exp(-ca/l)]. \quad (7)$$



Slip correction for motion in gases

Here l is the mean free path of the gas molecules, a the sphere radius and A , B and c are constants. The slip factor F multiplies the Stokes' law velocity to yield the true value. Their upper limit of l/a was 188.

Millikan (1920) published his final values for these constants after experimental work on the electronic charge which extended over ten years. His range of l/a was from 0.5 to 134 and the experiments were described in 1923 (b). His value for $A+B$ agreed with that of Knudsen and Weber within 1%, although derived from work with minute oil drops falling freely through air. When the pressure is low ($l/a > 20$) the slip factor becomes equal to

$$1 + \frac{l}{a} (A + B).$$

Millikan considered (1923 a), from a review of capillary-tube experiments by Warburg and by Knudsen, that the coefficient of slip between glass and air was exceedingly close to that between oil and air. At higher pressures, although the total correction diminishes, the exponential term becomes important; under these conditions Knudsen and Weber found a smaller slip correction than Millikan, that is better agreement with Stokes' law, and Millikan suggested this might be due to wall effect.

Mattauch (1925) carried out work on rather similar lines to Millikan, using oil drops in nitrogen. His experiments extended only up to $l/a=5$ and down to 0.1. In the region immediately adjoining that in which Stokes' law is valid, the slip factor reduces to

$$l + A l/a.$$

Monch (1933) experimented with tobacco smoke, in this region only, and found the same value for A as did Mattauch.

In table 7 the constants determined by these observers are summarized. They have been worked out for the following definition of the mean free path:

$$l = \frac{\eta}{0.3502\sigma\bar{c}}.$$

In the figure the function $A + B \exp(-ca/l)$ is plotted for the constants of each experimenter.

Table 7

Author	A	B	c	$A+B$
Knudsen and Weber	0.772	0.400	1.63	1.172
Millikan	0.864	0.29	1.25	1.154
Mattauch	0.898	0.312	2.37	1.210
Mönch	0.90	—	—	—

The derivation of a working equation from these results will now be discussed. Consider first the constant A ; this is important at low values of l/a . There is agreement within 4% between Millikan, Mattauch and Monch, but Knudsen and Weber are very low. Millikan, as mentioned above, gave grounds for suspecting their figure. In addition (1923 a) he derived a relationship between the constant A and a slip coefficient which could be measured by rotating-cylinder experiments, using hydrodynamic theory, at pressures where slip was small. Such experiments, by his associates, provide a good check on his oil-drop figures. Next, there exists no confirmation that the slip coefficient for tobacco-smoke particles in air is the same as for oil in air or nitrogen, which is assumed in comparing the value of Monch with those of the others. Thus it seems advisable to average the values of A given by Millikan, Mattauch and Monch, giving double weight to that of the first-named. Knudsen and Weber's figure is rejected. This gives us $A=0.882$.

For high values of l/a we have only the data of Millikan and of Knudsen and Weber to consider, since Mattauch's observations do not extend beyond $l/a=5$. These agree very well, and the average of their values of $A+B$ is 1.163, giving $B=0.281$.

It now remains to assign a value to the constant c , which is operative for intermediate values of l/a . A single point on the graph will serve to define this, and the point (2,1.010) has been selected, this being the mean of the values of Millikan and Mattauach, approximately in the middle of the latter's range of l/a . Calculation then gives $c = 1.57$.

Hence the following equation for the slip factor has been obtained :

$$F = 1 + l/a[0.882 + 0.281 \exp(-1.57 a/l)]. \quad \dots\dots(8)$$

The relevant part of this equation is shown on the graph for comparison with the figures of the workers from the results of whose labours it has been deduced.

If the latest definition of the mean free path, due to Chapman and Enskog, is used,

$$l = \frac{\eta}{0.499\sigma\bar{c}};$$

this transforms to

$$F = 1 + (l/a)[1.257 + 0.400 \exp(-1.10 a/l)]. \quad \dots\dots(9)$$

For experiments in air, equation (8) can be put into a convenient form free from kinetic theory constants, since

$$pl = 7.16 \cdot 10^{-4},$$

where p is the pressure in cm. mercury.

$$F = 1 + (10^{-4}/pa)[6.32 + 2.01 \exp(-2190 pa)]. \quad \dots\dots(10)$$

In using these equations it must be remembered that differences in the constants may occur owing to the conditions of reflection of gas molecules being governed by their own nature and by that of the surface of the sphere. Millikan (1923 a) gives a list of values of A for various interfaces and Epstein (1924) discusses the question theoretically.

ACKNOWLEDGMENT

The author wishes to express his appreciation of the encouragement and advice which he has received from Dr. T. Bedford.

REFERENCES

- ALLEN, H. S., 1900. *Phil. Mag.* 50, 323, 519.
 ARNOLD, H. D., 1911. *Phil. Mag.* 22, 755.
 BUCKINGHAM, E., 1914. *Phys. Rev.* 4, 345.
 CASTLEMAN, R. A., 1926. *N.A.C.A. Tech. Note*, no. 231.
 DAVIES, C. N., 1939. *Thesis and Dissertation* (London).
 EPSTEIN, P. S., 1924. *Phys. Rev.* 23, 710.
 KNUDSEN, M. and WEBER, S., 1911. *Ann Phys., Lpz.*, 36, 982.
 LEMIN, C. E., 1931. *Phil. Mag.* 12, 589.
 LIEBSTER, H., 1927. *Ann. Phys., Lpz.*, 82, 541.
 LUNNON, R. G., 1928. *Proc. Roy. Soc. A*, 118, 680.
 MATTAUCH, J., 1925. *Z. Phys.* 32, 439.
 MILLIKAN R. A., 1920. *Phys. Rev.* 15, 545 ; 1923 a *Ibid.* 21, 217, 223 , 1923 b. *Ibid.* 22, 1.
 MÖLLER, W., 1938. *Phys. Z.* 39, 57.
 MÖNCH, G., 1933. *Phys. Z.* 34, 77.
 SCHILLER, L., 1932. *Handb. Exp. Phys.* 4, 2 T. 339.
 SCHMIEDEL, J., 1928. *Phys. Z.* 29, 593.
 WIESELSBERGER, C., 1922. *Phys. Z.* 23, 219.
 WIESELSBERGER, C., 1923. *Ergeb. Aerodynam. Versuchsanstalt, Göttingen*, 2, 28.

A PHYSICAL THEORY OF THE SOLAR CORONA *

By MEGHNAD SAHA, F.R.S.,

University of Allahabad

Lecture delivered 23 November 1944 ; MS received 6 April 1945

§ 1. INTRODUCTION

NEARLY twenty-five years ago, when the present writer was preparing his paper "On a physical theory of stellar spectra" (Saha, 1921), he had the benefit of very sound advice from the late Professor Alfred Fowler, who allowed him to make free use of his (Fowler's) own unrivalled knowledge of spectroscopy and of stellar spectra. Fowler's remarks on this theory, which to my knowledge were never put in print, may now be disclosed. "The thermal ionization theory", he told me repeatedly, "accounts in a general way for the spectra of normal stars; but there are very important exceptions, e.g. the stars with peculiar spectra, the planetary nebulae; even in the case of normal stars, the great strength of Balmer lines of hydrogen which persists throughout all stellar classes is a disquieting feature, and in the case of the sun, the peculiar behaviour of helium cannot, in my opinion, be accounted for by the thermal ionization theory at all".

During the past twenty-five years, many of these points raised by Fowler have been taken up by well-known workers: Darwin, R. H. Fowler and Milne, Zanstra, and others in this country, mostly on the theoretical side; and by Russell, Bowen, Struve, Menzel, Payne, and their co-workers in the U.S.A., Unsöld, Pannekoek, and other workers on the Continent. But the helium problem appears to have remained very much as it was twenty-five years ago. Briefly the problem is as follows: The Fraunhofer spectrum of the sun shows only the lines of such elements as have excitation potentials (energy values of the lower state) between zero and 10 volts; in the chromospheric spectrum, the lines of ionized elements are relatively stronger, but in no case, helium excepted, do we get lines of stronger excitation than 14 to 15 volts (energy value of upper state). The lines of He do not occur at all in the normal Fraunhofer spectrum, except over disturbed regions, like penumbra of sunspots, but occur prominently in the flash spectrum up to heights of 7500 km. These lines have an excitation potential exceeding 20 volts; but the line of ionized helium $\lambda 4686$, $\nu = 4R \left(\frac{1}{3^2} - \frac{1}{4^2} \right)$ occurs as a prominent but low-level chromospheric line scarcely exceeding 2000 km. in height. This line has an excitation potential of about 75 volts, and one fails to see how such high excitation can exist in the sun, and that too in the lower levels.

The points were repeatedly urged by Professor A. Fowler, and were repeated by myself later on many occasions, and have also received attention from others.

* For fuller details, see Saha (1942).

There are certain additional features regarding the occurrence of He lines. I think it was Evershed who first drew attention to the fact that the chromospheric He lines tend to get fainter and ultimately disappear towards the limb. The matter was confirmed by Pannekoek, and Minnaert (1928), and more fully by Perepelkin and Melnikov (1935). The findings of the later workers are represented in table 1 and figure 1, taken from their works.

Table 1

Height (km.)	E (erg/cm ² sec)	Height (km)	E (erg/cm ² sec.)
500	39×10^{-8}	4000	60×10^{-8}
1000	125	4500	29
1500	186	5000	12
2000	212	5500	4.3
2500	195	6000	1.3
3000	151	6500	0.3
3500	100	7000	0.1

These results are inexplicable on the ionization theory, or any modification of it. For some time past I have been thinking of another explanation, which I hesitated to put forward on account of its radically heterodox nature. Allowing that He exists in some quantity in the solar atmosphere, it is clear that neither the ultra-violet radiation from the sun nor the thermal conditions existing on the

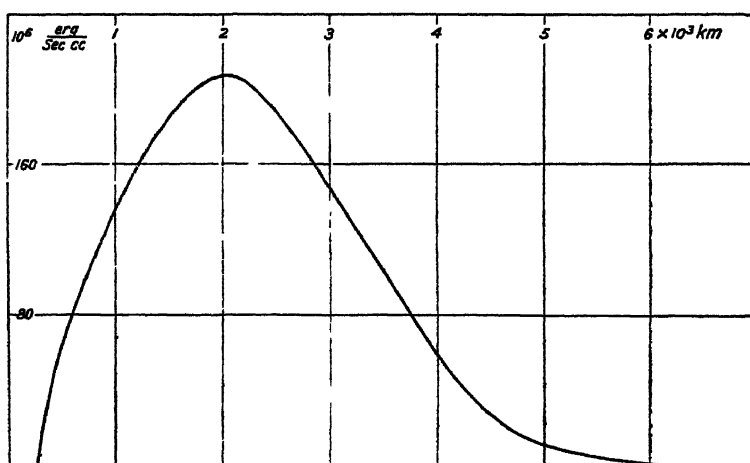


Figure 1

surface of the photosphere is capable of exciting it to luminescence in the way we obtain in the sun. The suggestion regarding their origin is as follows:— First, suppose that α -particles are constantly being produced throughout the solar surface, as a result of some nuclear reaction, and hurled forth through the solar atmosphere. As they pass through the solar gases (mostly hydrogen), they go on ionizing these atoms by collision (as in J. J. Thomson's theory of ionization by collision), and continuously losing energy. When their energy

has sufficiently diminished, they capture an electron in any orbit and become normal or excited He^+ . The excited He^+ atom may radiate energy of which only $\lambda 4686$, and possibly the lines $\nu = 4R \left(\frac{1}{4^2} - \frac{1}{m^2} \right)$, are within observable range.

The He^+ atom moves forward along the original direction but it goes on losing energy, which is spent, as in the case of He^{++} , in releasing electrons from atoms by collision. When its velocity of motion has sufficiently diminished, it may capture a second electron, and become a normal or excited He atom. The excited He atom gives us the high-level chromospheric He lines.

This phenomenon of capture of electrons by α -particles to form He^+ and He was discovered by Rutherford and Henderson (1923) while studying α -tracks in the cloud-chamber. The capture of the first electron begins to take place when the velocity of the α -particle has fallen to $2c\alpha$, where c is the velocity of light and α the Sommerfeld constant. It may be recalled that $c\alpha$ is the velocity of the electron moving in the first orbit of the H atom. We have $c\alpha = 2.18 \times 10^5 \text{ km./sec.}$ We shall have frequently to express velocities in this paper in terms of $c\alpha$ as unit in the form $V = sc\alpha$, where s is a numerical coefficient. In the cloud-chamber, when the α -particle starts to move, it does so with a velocity of the order of $9c\alpha$ (for α -particles of range 11 cm. from Th C''). It goes on producing electrons by collision, and thus gradually loses energy. When the last centimetre is reached, and its velocity has reached $2c\alpha$, corresponding to an energy of 1×10^5 volts, and range of about 0.46 cm., the α -particle begins to capture electrons to an appreciable degree.

But He^+ which is formed by the capture of an electron may again lose this electron by collision with atoms, and again become He^{++} or α -particle. In fact Rutherford (1924) showed that this alternate loss and capture of electrons may occur thousands of times within the last millimetre of the range of the α -particle, but all the time the velocity of the α -particle or of He^+ is falling, and when it reaches $\simeq c\alpha$, He^+ may capture a second electron from cloud-chamber gases and become He. But this may be again ionized to He^+ , until ultimately we get He, and the track terminates.

A mathematical theory of this effect has been worked out by Oppenheimer (1928) and by Kramers and Brinkmann (1930), and applied by Jacobsen (1935) for explaining the velocity-range phenomenon in the experiments of Rutherford and Henderson, and also in his own experiments.

The suggestion regarding the occurrence of He^+ and He lines is equivalent to saying that the cloud-chamber phenomena described here occur on the sun on a gigantic scale, but the α -particles are due not to natural radioactive bodies but to some reaction taking place on the solar surface. In the cloud-chamber, some of the α -particles must be capturing electrons in excited orbits, but we cannot observe emission of the characteristic lines of He, owing to their feebleness. The same is true of the capture of electrons from atoms by He^+ . But in the sun the captures are sufficiently numerous and the lines emitted are strong enough to be observed in the flash. The explanation accounts in a satisfactory manner, at least qualitatively, for the occurrence of $\lambda 4686$ in some strength in the lower chromosphere (up to a height of 2000 km.), and of the occurrence of He lines in the higher chromosphere up to heights of 7500 km., and also their tendency to disappear towards the limb.

There is only one apparent difficulty in this hypothesis of the origin of He^+ and He lines on the sun. According to laboratory observations so far available, the capture of the electron by He^{++} begins to take place when $V \simeq 2c\alpha$, and of the second electron by He^+ when $V \simeq c\alpha$. The He atoms in the sun ought therefore to be in motion with velocities of this order. But this is not apparently observed, though the He lines are actually found broad. The explanation is probably to be found in the fact that α -particles originate below the reversing layer, and by the time they come out of this region they have dissipated most of their energy in the process of ionizing other particles by collision.

If these suggestions stand criticism, it should be possible for us to calculate the intensity of ultra-violet emission from the sun due to He^+ and He, and estimate their relative importance in promoting ionization of the earth's upper atmosphere.

Alpha-particles are produced in many nuclear reactions, and at this stage it is needless to look for any particular reaction which may be mainly responsible for its production on the solar surface. The question is whether α -particles on such vast scales can be produced on the surface of the sun. If so, what is the subsequent fate of these particles? Do they sink deep, get doubly ionized in the interior, and contribute to restoring the α -particle balance of the interior of the sun? These questions may stand for the moment.

It is also worthy of notice that though the visible lines of He are not usually found in the Fraunhofer spectrum, Babcock (1934) records $\lambda 10830$, which is $1s2s^3S - 1s2p^3P$ as a faint absorption line in the infra-red part of the Fraunhofer spectrum. This line requires for its production as absorption line some accumulation of He in the $1s2s$ state, which is metastable. This indicates that He exists in some strength in the reversing layer in the normal $1s^2$ and $1s2s$ states, but not in the $1s2p$ or any higher state. The finding is not, in my opinion, antagonistic to the hypothesis of formation of He in the solar atmosphere out of α -particles.

It is obvious that the hydrogen atmosphere of the sun may also originate, at least partly, in the same way, for the proton is also a most frequent product of nuclear reactions. But a hydrogen atom once formed by the capture of an electron by the proton in the first or, better, in the second orbit can be sustained by radiation pressure, so its career should be fundamentally different from that of the He atom.

It is the belief of the present author that many outstanding problems of the solar and stellar atmospheres, such as prominences, spots, flares giving rise to radio fade-outs, may find their explanation in nuclear reactions taking place more vigorously on limited parts of the surface. It is quite probable that nuclear reactions of the type considered take place more vigorously in the interior, as shown by Bethe (1939) and Gamow (1939), but the probability of their occurrence on the surface on a reduced scale cannot be excluded. For example, it has been found that the He line $\lambda 5876$ occurs as an absorption line in the neighbourhood of disturbed areas, namely, penumbrae of spots. Probably nuclear reactions producing particles are the cause of formation of such disturbed regions, and reactions are much more vigorous than on the normal surface, and a temporary He atmosphere sufficient to give us D_3 in absorption may be formed in these regions.

§ 2. THE PROBLEM OF THE SOLAR CORONA

Extraordinary interest, in spite of the war, has been aroused in recent years in the problems of the outermost part of the solar atmosphere (inner and outer corona) by the work of Edlén (1942) on the identification of coronium lines. The story of this identification has been told by Russell (1941), by Swings (1943), and by Edlén himself in an exhaustive memoir (1942), and need not be repeated here. It appears to have been conclusively proved that most of the coronal lines are due to atoms of Fe, Ni and Ca which have lost a large number of their outer electrons, sometimes amounting to as many as fifteen or sixteen. The details of this identification, as far as required for our purpose, are given in table 2.

Table 2*

Wave-length (A.)	Intensity		Identification		Ionization potential (ev.)
	Grotrian	Lyot	Ion	Transition	
3328.1	1 0		Ca XII	$^2P_{3/2}-^2P_{1/2}$	589
3388.10	16		Fe XIII	$^3P_2-^1D_2$	325
3454.13	2.3				
3600.97	2.1		Ni XVI	$^2P_{1/2}-^2P_{3/2}$	455
3642.87			Ni XIII	$^3P_1-^1D_2$	350
3800.77					
3986.88	0.7		Fe XI	$^3P_1-^1D_2$	261
4086.29	1.0		Ca XIII	$^3P_2-^3P_1$	655
4231.4	2.6		Ni XII	$^2P_{3/2}-^2P_{1/2}$	318
4311.5					
4359					
4567	1.1				
5116.03	4.3	2.6	Ni XIII	$^3P_2-^3P_1$	350
5302.86	100	120	Fe XIV	$^2P_{1/2}-^2P_{3/2}$	355
5694.42		1.5			
6374.51	8.1	28	Fe X	$^2P_{3/2}-^2P_{1/2}$	233
6701.83	5.4	3.3	Ni XV	$^3P_0-^3P_1$	422
7059.62		4			
7891.94		29	Fe XI	$^3P_2-^3P_1$	261
8024.21		1.3	Ni XV	$^3P_1-^3P_2$	422
10746.80		240	Fe XIII	$^3P_0-^3P_1$	325
10797.95		150	Fe XIII	$^3P_1-^3P_2$	325

* Taken from an article by Swings (1943).

As there appears to be no way of denying the accuracy of the identification, the astrophysicist is faced with a number of problems of a unique type, which may be enumerated as follows :

- (1) What is the physical process giving rise to such highly charged ions ?
- (2) How can these highly charged ions, once produced, maintain their charge in the solar atmosphere ?
- (3) To explain the other characteristics of these lines noted by Lyot (1939) and in the eclipse expeditions, namely, the great breadth of these lines towards their base, sometimes amounting to 1 A., which gradually diminishes outwards, the intensity variations of these lines, etc., along with phases of solar activity.

These may be called "Coronium problems", in contrast to the second set of problems now to be discussed which may be called the "Corona problems". The two sets of problems must be discussed together as they are complementary. The coronium lines are found to occur in the "Inner corona"—which extends from beyond the top of the chromosphere (height, 14,000 km.), sometimes to a distance of about 10^7 (4×10^5 km.) from the photosphere. The inner corona shows besides the coronium lines, a continuous spectrum, which, though nearly a million times fainter, is of the same type as the photospheric spectrum, but with the Fraunhofer lines blurred out. In the outer corona, however, the coronium lines disappear, but the Fraunhofer lines reappear in its continuous spectrum.

The continuous spectrum of the corona has received attention from a number of workers, namely, Minnaert (1930), Grotrian (1933, 1934), and several others. They have proved that it is due to the Rayleigh scattering of photospheric light by an atmosphere of electrons as suggested by Schwarzschild nearly thirty years ago. From the variation of intensity of the coronal light with distance from the photosphere, it is possible to estimate the electron density at different heights, and the figures for a mean corona are reproduced in table 3.

Table 3 Electron density at various heights in the corona*

h (minutes of arc)	N	h (minutes of arc)	N
0.00	4.58×10^8	22.4	1.79×10^6
0.48	3.11	25.6	1.35
0.96	2.29	28.8	1.10
1.6	1.56	32.0	9.13×10^5
3.2	7.04×10^7	40.0	6.32
4.8	3.84	48.0	5.12
6.4	2.38	64.0	3.81
9.6	1.11	80.0	2.49
12.8	6.13×10^6	112.0	1.63
16.0	3.73	144.0	1.10
19.2	2.50		

* Taken from Unsold's *Sternatmosphäre*, 1939, chap. 17

The great difficulty has been to find the source of the electrons constituting the corona. They cannot arise from thermal or photoelectric ionization of solar atoms, as we have then to postulate in coronal heights the existence of a comparable concentration of atoms and ions, which is impossible on dynamical grounds. The best hypothesis appears to be that of Minnaert (1930), and may be given in his own words: "Anderson (1926) has shown that the corona cannot be in equilibrium if the ordinary physical laws are valid. Instead of assuming, as he does, that very hypothetical laws must be applied, we may attempt to account for the corona by assuming that it really is not in equilibrium, and that its particles are continuously being projected towards space."

According to Grotrian (1934), the continuous spectrum of the scattered radiation from the *inner corona* shows depressions in regions corresponding to chief Fraunhofer absorption lines, but amounting in width to about 100 Å, but the lines

reappear in the outer corona. He sought to explain the first observation by the hypothesis that the electrons in the inner corona are moving outwards with velocities of the order of 4000 km./sec. According to Moore (1934), the velocity of the coronal streamers (electrons) amounts to 20 to 30 km./sec. These figures probably refer to the outer corona.

Many investigators, including Edlén himself, have sought to explain the occurrence of coronium lines on a temperature basis. The arguments are two-fold :

(1) The coronal lines are, according to measurements of Lyot, quite broad, of the order of 1 Å., and if the width be due to Maxwellian motion of the emitting particles, the temperature ought to be 2.34×10^6 ° c. This temperature is sufficient to produce the required amount of ionization of the Fe and other atoms.

It is, however, difficult to think of any physical mechanism by which such high temperatures can be produced all over the outer layers of the sun. "Temperature" always means some equilibrium condition, and possibly a small black-body, placed at the coronal heights, would not show a higher temperature than 3000 to 4000° c. We may, however, have high local temperatures over limited regions, as in the case of a rocket burst in our own atmosphere, where we may have a small region around the rocket in which very high temperatures prevail for a short period of time. Several workers have hinted that the production of highly stripped iron, nickel and calcium ions responsible for the emission of coronium lines may be due to the bombardment of the solar atmosphere by meteoric matter in the way imagined by Lindemann and Dobson for explaining meteoric flashes in the earth's atmosphere. But the essence of the Lindemann-Dobson theory is that the meteor, striking the earth's atmosphere with a velocity ranging between 7 and 26 km., drives before it the whole column of air in its path, which is heated by adiabatic compression to a temperature sufficient to bring meteoric matter to luminescence. Lindemann and Dobson found that the gas pressure at the heights where the meteor strikes should be far larger than could be concluded from meteorological considerations. In the sun, the meteoric matter would fall with a velocity of 622 km./sec., but would probably get vaporized long before it reached the chromosphere, and even if some fragments escaped vaporization, the amount of matter in its path would be far too small for production of high temperatures according to the Lindemann-Dobson process. For the meteoric matter which vaporized, the atoms would be rushing with velocities of the order of 622 km./sec. into the solar atmosphere. The effect on these atoms may be obtained by supposing them to remain at rest, and allowing the solar atoms to rush past them with velocities of the order of 622 km./sec. As far as free and bound electrons are concerned, this is equivalent to bombarding the atoms with electrons having an energy of ≈ 10 volts, which is not sufficient even to tear the outermost electrons from the meteoric atom. As far as the nuclei of solar atoms are concerned, we need take only H nuclei. Their energy is of the order of 5000 volts, and such particles can tear out only one or two outer electrons at each encounter. The meteoric atoms can be deprived of 10 to 14 electrons only when they plunge very deep into the sun, but not at coronal heights. Further, Waldmeier (1938) has shown that the contour of the width curve of the coronal lines as found by Lyot can also be explained on the supposition that the emitters

of coronal lines are streaming outwards or inwards with a velocity of 60 km./sec. Lyot (1934) has further found that the width is largest when the emitters are nearest the sun's limb, and becomes narrower as the height increases. This, combined with Waldmeier's suggestion, shows that the emitters of coronal lines are streaming out of the sun with velocities which go on diminishing as greater heights are reached. The meteoric flash theory is not therefore sufficient to explain either the high ionization or the increasing width of coronal lines towards the solar limb, as actually found by Lyot and Waldmeier (1944).

The writer (Saha, 1942) has ventured to suggest that the Fe and other ions responsible for the emission of the coronium lines are due to some nuclear process

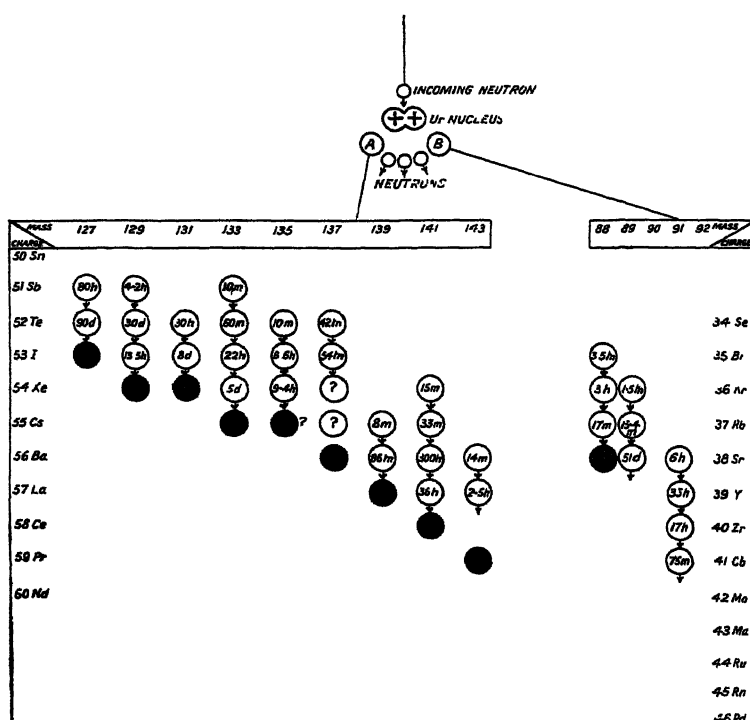
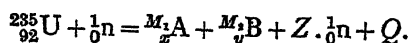


Figure 2

identical with or akin to that of nuclear fission, discovered by Hahn and Strassmann in 1939. The story of this discovery may be read in several excellent reports on the subject (Livingston, 1941; Walke, 1941), but the facts necessary for astrophysical purpose may be briefly described. It was found by Hahn and Strassmann that when heavy nuclei like ^{238}U , ^{235}U , ^{232}Th , ^{239}Pa are bombarded by neutrons, fast or slow, they break up according to the scheme (for ^{235}U) (figure 2):

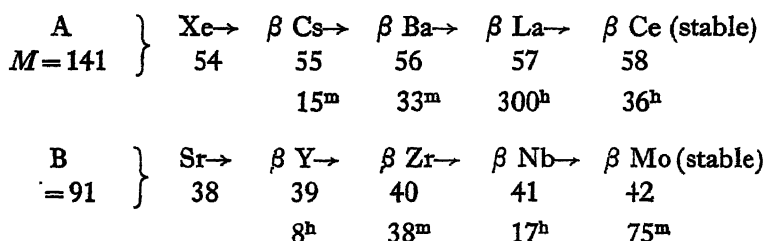


$M_x\text{A}$ and $M_y\text{B}$ are nuclei having respectively the charge numbers x and y , and mass-numbers M_1 , M_2 ; Z is the number of neutrons, generally 3 or 4, which evaporate in the process (Szilard and Zinn, 1942); Q is the energy evolved in the process.

$$x + y = 92, \quad M_1 + M_2 = 236 - Z.$$

The A products have been found to have x varying from 46 to 60, the B products have y varying from 35 to 46. The reaction is exothermic, as can be seen from mass relationships, and Q is $\simeq 200$ mev. for binary fission, and it is distributed, according to the law of conservation of energy and momentum, between A and B, A receiving $QM_2/(M_1 + M_2) = Q_A$ and B receiving $QM_1/(M_1 + M_2) = Q_B$, respectively.

Neither the A nor the B products are stable on account of the high proportion of neutrons, but each has to emit 3 or 4 β -rays successively, till they are reduced to stable forms, as might be illustrated in the chain processes



The fission process is beautifully illustrated in the Wilson-chamber photographs taken by Corson and Thornton (1939), Böggild *et al.* (1941) in Prof. Bohr's laboratory before Denmark was invaded.

What is important for our purpose is the high energy with which the fission fragments are thrown out in the reaction. To take an example : If $x=54$, $M_1=141$, $y=38$, $M_2=91$, $Z=4$, we have $Q_A \simeq 80$ mev and $Q_B \simeq 120$ mev. The velocities corresponding to these energies are $V_A \simeq 4.7 c\alpha$, $V_B \simeq 7.1 c\alpha$. These velocities are much larger than the orbital velocity not only of the outer electrons of the stable products, but also of many of their inner electrons. Bohr (1941), Knipps and Teller (1941) and Lamb (1941) have pointed out that as soon as the fission fragments are produced in any medium, they lose most of their outer electrons and can retain only those of their inner electrons whose orbital velocities are larger than, or comparable to, their own velocity. The fission particles therefore start as heavily ionized ones bereft of a large number of their outer electrons, and Bohr and Wheeler (1939) quote a Russian worker, Perlov (?), as having experimentally proved that the charge may be as high as 20.

The presence of Fe or Ni amongst the fission fragments has not yet been reported with definiteness, but considerations of energy and probability do not rule them out. Nishina and his co-workers (1939) have shown that the probability of a symmetric fission in which one fragment is much larger than the other increases with the energy of the bombarding neutrons. Further, ternary and quarternary fission are allowed by considerations of energy and probability, and Q in some cases may be as high as 250 mev., which is larger than the 200 mev maximum energy evolved in the binary fission process. It may be supposed that as a result of either of such processes taking place in the sun, smaller fragments are produced which, after a number of β -emissions, ultimately stabilize as nuclei of elements from Ca to Ni and are emitted with energies of the order of 60 mev. It is gratifying to note that after the writer had postulated such a process for the origin of coronal emitters, ternary and quarternary fissions were reported by Lark-Horowitz (1941) and theoretically treated by Present (1941).

It can be shown that if one of the heavy fission-elements undergoes a ternary or a quaternary fission, the fragments, after a number of β -emissions, will be elements from Ca to Ni (the limit on both sides is rather elastic), and they will be emitted with an energy of approximately 60 mev., i.e., in the case of Fe atoms with a velocity of 6.4 α . Let us now turn to table 4, which shows the velocity

Table 4* Stripped iron ions and their electron structure, etc

Ion	Electron structure	Funda- mental state	Value of the lowest terms in volts	$\sqrt{\frac{IP}{13.54}} = \frac{z_2}{n}$	Remarks.
²⁶ Fe I	3d ⁶ 4s ²	⁵ D ₄	7.83	0.76	Forbidden lines found in η Carinae. Bowen (1936).
Fe II	3d ⁶ 4s	⁶ D _{9/2}	16.5	1.10	
Fe III	3d ⁵	⁶ S	30.48	1.50	Bowen (1940) gives metastable lines found in nebulae. D Kundu thinks that some of these lines may occur in the corona.
Fe IV	3d ⁴	⁵ D ₀	56.8	2.05	
Fe V				(2.37)†	
Fe VI	..3d ³	⁴ F _{3/2}	..	(2.69)	" " " " " " ² D _{3/2} - ² D _{5/2} = 1875 cm ⁻¹ No metastable line available.
Fe VII	..3d ²	³ F	..	(3.01)	
Fe VIII	..3d	² D	150.4	3.33	
Fe IX	..3p ⁶	¹ S ₀	233.5	4.15	No metastable state.
Fe X	..3p ⁵	² P	261	4.39	$\lambda 6374.75$ ² P _{3/2} ← ² P _{1/2} .
Fe XI	..3p ⁴	³ P	288.9	4.62	$\lambda 7892$.
Fe XII	..3p ³	⁴ S	320	(4.91)	Has no metastable line in the available range
Fe XIII	..3p ²	³ P ₀₁₂	346	(5.06)	{ 10746.80. 10797.95 $\lambda 5303$ ² P _{1/2} - ² P _{3/2} No metastable state.
Fe XIV	..3p	² P	373	5.25	
Fe XV	..3s ²	¹ S ₀	454	5.79	
Fe XVI	..3s	² S _{1/2}	487	5.99	" "
Fe XVII	..2p ⁶	¹ S ₀	1259.7	9.65	" "

* Reproduced from the author's paper, "On a physical theory of the solar corona", *Proc. Nat. Inst. Sci.* 8, 99 (1942).

† Parentheses () denote that the value is extrapolated

of the outermost electrons of iron, and its ions. We find that the 2 *p*-electron of Fe XVII has a velocity of 9.6 α . This will therefore be retained, but if we take Fe XVI, the 3*s*-electron is found to have a velocity of 6 α . We can therefore conclude that in a fission process of the type envisaged here, occurring in the reversing layer, the iron atoms which normally have the electron composition 1*s*² 2*s*² 2*p*⁶ 3*s*² 3*p*⁶ 3*d*⁶ 4*s*² will have lost the outer 15 electrons, namely, 3*s* 3*p*⁶ 3*d*⁶ 4*s*², and will start as Fe XVI with the electron composition 1*s*² 2*s*² 2*p*⁶ 3*s*. It will now be interesting to follow the physical processes to which such a highly charged ion produced anywhere in the sun can give rise, as it passes through the solar atmosphere. These are :

(a) *Ionization by collision.* The ion goes on knocking electrons and nuclei from the atoms which it encounters in its way, just as an α -particle does when it is projected in a cloud-chamber. In this process, the ion continuously loses energy, and it is possible to calculate its *range* with the aid of the Bohr-Bethe formula (1936), provided we assume that the solar atmosphere consists mainly of H atoms with few C and N atoms (as given by Menzel). The range is found on certain plausible assumptions to be $R_H \simeq 1.31 \times 10^{21}$ H atoms/cm² for Fe xvi projected with a velocity of $6.4 c\alpha$.

We can draw very important conclusions from this calculation. Nuclear processes giving rise to Fe xvi, or similar highly charged iron or other ions may occur throughout the whole *solar interior*, but most of such particles have no chance of ever passing out of the solar atmosphere. Most of them will stop dead earlier. If the number of H-particles in the reversing layer is taken to be 1.8×10^{22} /cm² (Unsold, 1939), R_H is far less than this number, and it is obvious that only those fission particles which are produced rather high up in the "reversing layer" have a chance of escaping through the chromosphere and emerging into the inner corona.

(b) *Possibility of the further loss of an electron by the ion.* The Fe ion may itself lose a further electron by collision with atoms, but the probability of this event vanishes when the velocity of the ion reaches a certain limiting value. For the solar atmosphere, the limiting velocity, under certain plausible assumptions, is $V_e \simeq c\alpha \cdot z$, where z is the net charge on the ion. The Fe xvi ion cannot therefore lose any further electron.

(c) *Capture of electrons by the ion.* The Fe xvi may capture an electron from any one of the atoms in its path, or even a free electron, and become Fe xv: $1s^2 2s^2 2p^6 3s^2$ (normal) or $3s nx$, where nx is a higher orbit. The probability of the capture at first increases as the velocity of the ion diminishes. When the capture is in an excited orbit, the iron is expected to execute one or more quantum transpositions, emitting x rays, and ultimately we shall have normal Fe xv. This has no metastable levels, so no visible radiation can be emitted by Fe xv.

The Fe xv ion will now pass through the same career as Fe xvi, but the electron composition of the next ion formed is $1s^2 2s^2 2p^6 3s^2 3p$, and hence we have two metastable levels, $^3P_{1/2}$ and $^2P_{3/2}$. The strong coronal line $\lambda 5303$ is due to the forbidden transition $^2P_{3/2} \rightarrow ^2P_{1/2}$. The emission of $\lambda 5303$ necessarily indicates that some amount of x radiation due to the allowed transition $3s^2 nx \rightarrow 3s^2 3p$, $\lambda \simeq 80\text{\AA}$, is also being emitted. The capture can take place at all velocities of Fe xv, from $s=6$ down to $s=0$, but the formula of Brinkmann and Kramers has been worked out only for capture in *S*-orbits and for high velocities of the ion. It has still to be worked out for small s and for capture in *p*-orbits; hence at this stage it is not possible to give any quantitative estimate. The next ions from Fe xiii to Fe x are all formed by successive capture of electrons from the solar atoms or of free electrons in the solar atmosphere, and thus the $3p^2$ -shell is formed ($x=1$ to 5), which gives us the coronal radiation given in table 2.

The possibility of any one of the Fe xiv to Fe x ions emerging out of the

chromosphere to the coronal heights therefore depends upon : (i) the probability of a fission of the type mentioned above taking place in the sun, (ii) the region where the fission takes place. Bohr and Wheeler (1940) have shown that only heavy nuclei like ^{238}U , ^{235}U , ^{232}Th , ^{229}Pa are capable of fission. These can occur in the solar interior as well as in the reversing layer, but ions formed in the interior are stopped dead earlier, and only such as originate in the reversing layer can escape to coronal heights. The origin of the coronium emitters is therefore to be found in the upper part of the reversing layer.

At this stage some of the probable doubts and objections in the mind of the reader may be anticipated .

(1) Have we in the sun's atmosphere or interior sufficient U or Th atoms, which alone have been shown by Bohr and Wheeler (1939) to be capable of fission by neutron-bombardment ?

U has not yet been traced in the sun, probably owing to the extreme complexity of its spectrum. But the presence of lines of singly-ionized Th has recently been definitely established by Moore and Babcock (1943).

(2) Why should we get only Fe, Ni and Ca in the corona, and not any of the contiguous elements, say Co, Mn, or Cu, A or K ?

An investigation carried out by D. Kundu (1942) at Calcutta shows that Fe, Ni and Ca are spectroscopically better suited for identification than any other elements in this group excepting Co. But Co is probably represented by a faint line in the corona, $\lambda 4359$, which Kundu attributes to Co xv. Subsequently, through the courtesy of the Astronomer Royal of England, the writer has been able to have access to a copy of Edlén's paper (1942) in which the same opinion has been expressed.

But even if Co, Mn and other elements of the group are subsequently found to be represented by some of the fainter coronal lines not yet identified, it is clear that these elements are represented far less strongly than Fe, Ni and Co. Edlén has tried to connect the phenomenon with the so-called cosmic frequency of elements. But probably the real reason is that in a nuclear process there is a greater probability of the occurrence of even-numbered atomic elements than of odd-numbered ones. Each one of the former is represented by four or more isotopes, but the latter generally (for example, Sc, V, Mn and Co) by a single isotope. There is, therefore, a greater chance of fission products ultimately transforming themselves after β -emissions to Fe, Ni and Ca than to Co, Mn and Sc.

(3) Why do we not observe the forbidden lines of Fe ions from Fe ix to Fe ii, and of the corresponding ions of Ni, amongst the coronium-lines ?

Fe ix has no metastable state and Fe viii has the composition $1s^2 2s^2 2p^6 3s^2 3p^6 3d$. The $(3d^2 D_{3/2} - 2D_{5/2})$ separation is too small to give a line in the visible range. Some of the forbidden lines of Fe vii to Fe iii having the composition of $3d^2$ have been traced by Bowen in the nebulae, and of Fe ii in η Carinae, and by Merrill (1943) in BF Cygni and other stars; Kundu thinks, too, that some forbidden lines of Fe v can be identified with fainter, doubtful lines in the solar corona. But these doubtful lines require further investigation, both as regards wave-length measurements and identification.

It is clear that even if subsequent investigations prove that the forbidden lines of the $3d^x$ -complex (Fe VIII to Fe II) occur in the corona, they would prove to be extremely faint compared to the forbidden lines of the $3p^x$ combination. This may be due partly to the fact that the probability of capture of an electron in a d -orbit is far smaller than that in a p -orbit. Exact calculations are difficult and are being carried out, but the finding is questionable.

The complete establishment of these ideas will require a colossal amount of experimental and theoretical investigations, the nature of which is clearly indicated in the text. But the value of the hypothesis can also be assessed from a discussion of its bearing on the associated problems of the solar corona mentioned earlier.

If the considerations presented here regarding the origin of emitters of coronium lines prove to be correct, it is obvious that the electrons constituting the inner and outer corona are simply the δ -rays liberated by the coronium-emitters (Fe XIV and others) from H and other atoms in the upper reversing layer, and the chromosphere in the process of ionization by collision as these highly charged emitters of coronium lines pass through the solar atmosphere. The velocity of these electrons is given by the relation $V_e = 2V_i \cos \phi$, where V is the velocity of the ion and ϕ the angle between the direction of emission of the δ -electron with the original direction of motion of the ion. It is clear that $V_e \simeq 2V_i \simeq 2c\alpha$, and may have as high values as $2c\alpha$. The swifter electrons are mostly emitted inside the reversing layer and inside the chromosphere, and will be able to escape with velocities of the order $2c\alpha$; they probably constitute the electron atmosphere which we know as the corona. The theory has evidently to be further worked out to yield more details about the corona.

The coronal problems are almost unique in astrophysics, because if we leave out the sun, the coronal lines are not observed in the very wide range of astrophysical phenomena; neither in spectra of normal or peculiar stars, nor in those of novae or supernovae, except in the solitary case of recurrent novae (RS Ophiuchi), as was discovered by Adams and Joy (1933), and confirmed by Swings and Struve (1943). But it is inconceivable to think that the sun should be a solitary exception. Probably the same physical processes which give rise to coronal lines are occurring everywhere, but the scale, compared with those of ordinary stellar emission, is far too low for the lines to be observable. We are able to observe them on the sun merely on account of our proximity to this star, and that only on special occasions (time of total eclipse) or by special devices (Lyot).

Is it possible to give more definiteness to the question of the *scale* of coronal emission compared to those of ordinary photospheric or chromospheric emissions? The photosphere emits 1500 cal./sec. per cm^2 of its area, the chromospheric emissions in $\text{H}\alpha$ come to about $\frac{1}{50}$ of the corresponding photospheric emission in $\text{H}\alpha$, and according to estimates of Lyot, the coronal emission in the green line is of the order of 10^{-6} A. of the corresponding photospheric emission.

The ideas presented in this lecture may be compared with those of Roseland (1934):—

“Considering, for instance, the most familiar case, that of the sun, it is surprising how few theories are of such an obvious character as to deserve unreserved

applause It will probably be admitted generally that the interpretation of the origin of Fraunhofer lines is now so far advanced that a revision of fundamental principles may be unnecessary in this field. Proceeding a little further to the interpretation of spectroheliograms, the ground is already getting considerably more insecure. And when we proceed still further, we meet the enigmas of the sunspots, the prominences, the chromosphere, and the corona, none of which can at present be said to be understood, even in the most liberal interpretation of the term.

“The enigmatic character of these phenomena is not so much concerned with the generally admitted fact that we do not understand their common cause, which underlies solar activity as a whole. It is more that we do not know how to interpret the individual manifestations in an intelligible manner. We know of no simple mechanism at present according to which magnetic fields of the magnitude observed in sunspots could be generated. The motion of prominences is recognized as quite different from any motion which could be produced by the combined action of gravitation and electromagnetic forces on a mass of gas in a vacuum, and the agglomeration of matter in the corona surpasses by billions the amount to be expected on any simple hydrostatic theory. These various facts have stimulated speculation to the breaking point, it being even suggested that here we witness our recognized physical laws set at naught by nature herself. Although these speculations are not likely to be taken very seriously by the experienced physicist, they bring out forcibly the unsatisfactory state of solar theory today.”

Rosseland's view in 1934 was : “Chromosphere, corona and prominences would in that case form a complex of dynamic phenomena, the theory of which must be based on considerations of the expansive motion of matter moving away from the sun in a more or less radial direction. It does not follow, of course, that all matter in a streamer is moving with the same velocity.”

Rosseland concludes : “Though we definitely do not know the nature of these primary particles, the existence of which is indicated by general arguments, there are reasons to believe that they are electrically charged.”

Rosseland has considered the equilibrium (?) of the electrified atmosphere, but the physical factors introduced (for example, resistance to the motion of positively and negatively charged particles) are of a vague character. Probably the ideas introduced here will impart definiteness to these factors.

The idea of temperature-equilibrium can be applied to the photosphere and the reversing layer, and that, too, very approximately. The general chromospheric phenomena and other associated ones like prominences are probably partly due to temperature, partly to radiation pressure, and probably nuclear reactions giving rise to α -particles and protons play some part. The coronal phenomena are of a different type—arising from a process akin to or identical with fission, and they are just like rocket-bursts in our atmosphere. The three types of phenomena intermix and produce a complicated picture.

The author had the privilege of discussing the theory of the corona given here with Professor Dirac during a short visit to Cambridge. Dirac made the most interesting suggestion that the β -rays emitted by the fission products may turn out to be the high-energy electrons which are wanted for explaining auroral

phenomena. For it is well known that Stormer's theory of the aurora has not been able to explain why the zone of maximum frequency of the aurora is at a distance of 22° from the magnetic poles (see Hewson, 1937). This proves that the corpuscular rays responsible for the auroral phenomena cannot be photo-electrons, or even β -rays of moderate energy. They can be either β -rays of energy of the order of 5 to 10 mev., or α -particles, but the last possibility is generally ruled out on other grounds. The β -rays expected to be given out by fission products have the requisite energy, but there are other factors, and the problem may be left at this stage.

This article was prepared in course of the author's tour through England and U.S.A., in 1944-1945 on a Government of India Mission, and he had the pleasure of discussing its contents with many friends, to whom his grateful acknowledgments are due. He is particularly indebted to Dr. J. A. Fleming, director of the Geophysical Institute, Washington, D.C., and Mr. Allan Shapley for unstinted help in the final preparation of the manuscript.

REFERENCES

- ADAMS and JOY, 1933. "Spectrum of RS Ophiuchi." *Pub. Astr. Soc. Pacific*, 45, 301.
 BABCOCK, 1934. "He lines in the Sun." *Pub. Astr. Soc. Pacific*, 46, 132.
 BETHE, 1939. "Energy-production in stars." *Phys. Rev.* 33, 434.
 BOGGILD *et al.*, 1941. "Range and straggling of fission-particles." *Phys. Rev.* 59, 275.
 BOHR, 1940. "Scattering and stopping of fission fragments." *Phys. Rev.* 58, 659.
 BOHR and WHEELER, 1939. "Mechanism of nuclear fission." *Phys. Rev.* 56, 426.
 BRINKMAN and KRAMERS, 1930. "Capture of electrons by α -particles." *Proc. K. Akad. wet. Amst.* 33, 973.
 CORSON and THORNTON, 1939. "Disintegration of uranium." *Phys. Rev.* 55, 509.
 EDLÉN, 1942. "On the identification of coronal lines." *Z. Astrophys.* 22, 30.
 GROTRIAN, 1933. "Variation of intensity of coronal lines." *Z. Astrophys.* 7, 26.
 HEWSON, 1937. "A survey of the facts and the theories of the aurora." *Rev. Mod. Phys.* 9, 403.
 JACOBSEN, 1935. "Capture of electrons by swift α -particles." *Phil. Mag.* 10, 401.
 KNIPPS and TELLER, 1941. "Energy loss of heavy ions." *Phys. Rev.* 59, 659.
 LAMB, 1941. "Passage of uranium fragments through matter." *Phys. Rev.* 58, 596.
 LARK-HOROVITZ and SCHREIBER, 1941. "U fission with Li-D neutrons." *Phys. Rev.* 60, 156.
 LIVINGSTONE and BETHE, 1937. "Report on nuclear physics." *Rev. Mod. Phys.* 9,
 LYOT, 1939. "Study of the solar corona and prominences." *Mon. Not. Roy. Astr. Soc.* 99, 580.
 MERRILL, 1943. "The spectrum of F-Cygni." *Astrophys. J.* 98, 473.
 MERRILL, 1944. "Spectroscopic observations of α Persei, etc." *Astrophys. J.* 99, 481.
 MINNAERT, 1930. "Continuous spectrum of the corona." *Z. Astrophys.* 1, 209.
 MOORE and BABCOCK, 1943. "Thorium in the Sun." *Pub. Astr. Soc. Pacific*, 55, 22.
 NISHINA *et al.*, 1939. "Fission of thorium by neutrons." *Nature, Lond.*, 144, 547.
 OPPENHEIMER, 1928. "Quantum theory of capture of electrons." *Phys. Rev.* 31, 349.
 PANNEKOEK and MINNAERT, 1928. "Observations of solar lines, etc." *Proc. K. Akad. wet. Amst.* 13, .
 PEREPEL'KIN and MELNIKOV, 1935. *Publications of the Pulkova Observatory*, no. 122, 14.
 PRESENT, 1941. "Possibility of ternary fission." *Phys. Rev.* 59, 467.
 ROSSELAND, 1933. "On the theory of the chromosphere and the corona." *Proc. Oslo Akad.*
 RUSSELL, 1941. "A puzzle solved." *Scientific American*, 165, 70.
 RUTHERFORD, 1924. "Capture and loss of electrons by α -particles." *Phil. Mag.* 47, 277.
 SAHA, 1921. "On a physical theory of stellar spectra." *Proc. Roy. Soc. A*, 99, 135.
 SAHA, 1942. "On a physical theory of the solar corona." *Proc. Nat. Inst. Sci. India*, 8, 99.

- SWINGS, 1943 "Edlén's identification of the coronal lines, etc" *Astrophys. J.* 98, 116.
 SWINGS and STRUVE, 1943 *Astrophys. J.* 97, 204.
 TURNER, 1940 "Nuclear fission." *Rev. Mod. Phys.* 12, 1.
 UNSOLD, 1939. *Steinatosphere*.
 WALDMEIER, 1937. "On the significance of contours of corona-lines" *Z. Astrophys.* 15, 44.
 WALDMEIER, 1944. "Observations of corona before and after total solar eclipse, etc." *Z. Astrophys.* 20, 250.
 WATKE, 1940. *Rep Progr. Phys.* 6, 16

ON TRACING RAYS THROUGH AN OPTICAL SYSTEM

(Fifth paper)

By T. SMITH, F.R.S.

(Communication from the National Physical Laboratory)

MS. received 23 February 1945

ABSTRACT An iterative method of tracing rays through an optical system which is suitable for operation by a fully automatic recording machine is described. The rays may be axial or skew, and the surfaces of any rotationally symmetrical form suitable for optical working. The relation of this method to earlier schemes and the advantages to be gained by successive approximations are considered.

FROM time to time experiments have been made at the National Physical Laboratory on methods of tracing single rays and pencils of rays through a series of lenses. It has been assumed that a calculating machine would be employed instead of tables of logarithms, and for this and other reasons the methods have differed from those described in Conrady's *Applied Optics*, which, it is believed, are still used to a considerable extent in the optical industry. An examination of the four earlier papers bearing the same title as the present note (Smith, 1915, 1918, 1920, 1921) will show that much of the original structure has been retained (though the notation has changed), but that some modifications have tended consistently in a definite direction. For example, the first paper gives formulae for paraxial rays, general skew rays, and the foci of sagittal pencils, which exemplify applications of the matrix formula

$$\begin{pmatrix} B & D \\ A & C \end{pmatrix} = \begin{pmatrix} 1 & D_0 \\ . & 1 \end{pmatrix} \begin{pmatrix} 1 & . \\ A_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & D_1 \\ . & 1 \end{pmatrix} \begin{pmatrix} 1 & . \\ A_2 & 1 \end{pmatrix} \cdots \begin{pmatrix} 1 & . \\ A_n & 1 \end{pmatrix} \begin{pmatrix} 1 & D_n \\ . & 1 \end{pmatrix} \cdots \quad (1)$$

to a system of n refracting surfaces. In this equation A_p represents a power associated with surface p and $-D_p$ the equivalent distance between surfaces p and $p+1$, i.e. the distance measured along the ray between these surfaces divided by the refractive index of the corresponding medium. The A, B, C, D on the left are well-known functions which determine the properties of the complete optical train. In the second paper the foci of tangential pencils are obtained by a slight generalization of the factors representing single refractions. The factor $\begin{pmatrix} 1 & . \\ A_p & 1 \end{pmatrix}$ is replaced by $\begin{pmatrix} B_p & . \\ A_p & C_p \end{pmatrix}$, where $B_p C_p = 1$. It will be seen that in all

cases the identity

$$BC - AD = 1$$

is satisfied. These formulae are still employed, and it is difficult to see how they can be simplified. They may be varied (Smith, 1930) to an almost unlimited extent without losing their main properties, and they lend themselves admirably to the rapid solution of problems which must be frequently encountered in industrial practice.

In some applications the values to be inserted in the elementary matrix factors are not known initially; they will be known when a preliminary calculation, which is usually considered the ray-tracing proper, has been made.

It is in this preliminary stage that a definite line of evolution is apparent. In the original scheme the variables* defining the position of a ray were the angle ψ it made with the axis of symmetry of the instrument and the length h of the perpendicular to the ray from the centre of curvature of the refracting surface, which was assumed to be spherical. If μ is the refractive index and ϕ the angle between the ray and the normal, the equations used were

$$\begin{aligned}\mu h &= \mu' h', \\ \sin \phi &= hR, \\ \sin \phi' &= h'R, \\ \psi' &= \psi - \phi + \phi',\end{aligned}$$

where accents are applied to symbols relating to the refracted ray and R is the curvature of the surface. At the following surface the first perpendicular is $h' + a \sin \psi'$, where a is the distance along the axis between the two centres of curvature. As systems are usually specified by radii of curvature (r) and axial distances (t) between successive surfaces, a is calculated from

$$a = t - r + r'.$$

This computing scheme is not without merits. It is simple to understand and to operate. Some common superfluous operations, such as finding the point of intersection of a ray with the axis after each refraction, are avoided. But experience suggested that improvements could be made in two respects. To obtain the greatest accuracy in calculations when only a small number of significant figures is retained, and particularly to obtain reliable differences between a number of calculations of the same type on the same system, the quantities involved should not be of very diverse magnitudes. In general this condition will not be satisfied if the centres of curvature are taken as reference points. This objection applies generally to all schemes which make use of the centres of curvature. The first change was therefore to drop the perpendiculars from the pole of the surface instead of from the centre of curvature. This ensured that the lengths used to define the position of a ray were roughly of the same magnitude at all surfaces. To retain this advantage throughout it was necessary to avoid using a radius of curvature as a multiplying factor. Each

* The curvature of a surface which is convex towards the incident light is positive, distances along the axis in the direction in which light travels are also positive. Perpendiculars to rays are positive if the foot of the perpendicular is above the axis. The angle ψ is positive if an anti-clockwise rotation from the positive direction along the axis is needed to bring a vector into coincidence with the onward direction of the ray; ϕ is positive if an anti-clockwise rotation from the direction of the normal would bring a vector into coincidence with the ray.

surface was therefore specified by its curvature instead of its radius, from the optical viewpoint this is clearly the scientific choice.

The second improvement was to eliminate all angular measures. The equation already given for ψ' is in appearance very simple, but the appearance is misleading when the numerical operations implied are taken into account. Three references to trigonometrical tables are required at each refraction, and subsidiary calculations for interpolation are usually involved. The whole process is therefore rather slow, notwithstanding its simplicity. In appearance the improved scheme looks more complicated. It will be best appreciated if refraction is first considered at a surface where there is no aberration. In that case h , the length of the perpendicular to the ray from the pole of the surface, is unaltered by refraction, and the equations are

$$\begin{aligned}\sin \phi &= \sin \psi + hR, \\ \mu' \sin \phi' &= \mu \sin \phi, \\ \sin \psi' &= \sin \psi - \sin \phi + \sin \phi'.\end{aligned}$$

Owing to the occurrence of aberration, h and h' are not equal, and the last equation is not true. But the difference between h and h' is of aberrational magnitude, and the value of $\sin \psi'$ just suggested is subject to a small correction. The correct formulae then consist of the first two equations together with

$$\begin{aligned}h' - h &= \frac{h (\sin \phi - \sin \phi') (\sin \phi' + \sin \psi)}{\frac{1}{2}\sigma^2 + \kappa}, \\ \sin \psi' &= \sigma - (h' - h)R,\end{aligned}$$

where

$$\sigma = \sin \psi - \sin \phi + \sin \phi'$$

and

$$\kappa = \frac{1}{2}[(\cos \psi + \cos \phi + \cos \phi')^2 - 1].$$

The h for the next surface is equal to $h' + t \sin \psi'$.

Since $h' - h$ is of aberrational magnitude, the denominator $\frac{1}{2}\sigma^2 + \kappa$ need only be known to a small number of significant figures. It represents twice the ratio of the first order to the total aberration, and at the National Physical Laboratory critical tables are now * used to evaluate it. One gives cosines when the sines are known, and two others, $\frac{1}{2}\sigma^2$ and κ , from the known values of σ and the sum of the three cosines.

This way of calculating rays is strongly preferred to all others that have been tried by those members of the Laboratory staff who have done much work of this kind. Differences of path and some other useful quantities are obtained from the values of $h' - h$ with almost trivial additions to the work, so that much information is obtained from the trace of a single ray.

A few years ago it was realized that the process of elimination and simplification could be carried further. A self-checking scheme for computing both axial and skew rays automatically, through surfaces of any shape, was under consideration. It was supposed that by means of punched cards a machine might be given the initial coordinates of a ray and a specification of the system through which it was to be traced. The machine would then carry out a cycle of operations and print all useful results. The cycle would be repeated, either for

* The original tables are not critical.

the instrument as a whole or alternatively surface by surface, until all the quantities repeated themselves exactly, when the work would be completed and the machine would in the one case stop and in the other proceed to the next surface.

To those who have only used the traditional methods of ray tracing the most surprising feature of the new scheme may well be the disappearance of Snell's law of refraction. The angles ϕ and ϕ' made by the incident and emergent rays with the normal at the point of refraction do not appear either explicitly or in disguise. These are the unessential quantities which have now been eliminated.

The most important quantities are the coordinate, measured parallel to the axis, of the point of refraction, and the product of the refractive index and the cosine of the angle the ray makes with the axis. If the point of refraction is x, y, z and the modified direction cosines of the ray are ξ, η, ζ , the main quantities when the axis of z coincides with the axis of the instrument are x and ζ . The equation of each surface will normally be referred to its pole as origin, and the standard form adopted is

$$s = f(z),$$

where s is the subnormal. A_p and D_p are then given by

$$\left. \begin{aligned} A_p &= \frac{\zeta_p - \zeta_{p-1}}{s'_p} \\ -D_p &= \frac{t_p - x_p + x_{p+1}}{\zeta_p} \end{aligned} \right\} \dots\dots (2)$$

If correct values are inserted for all the x 's and ζ 's, the analytical form of the law of refraction,

$$\frac{\xi_p - \xi_{p-1}}{x_p} = \frac{\eta_p - \eta_{p-1}}{y_p} = - \frac{\zeta_p - \zeta_{p-1}}{s_p},$$

shows that, if $A_{0,p}, B_{0,p}, C_{0,p}, D_{0,p}$ are the values of A, B, C, D obtained by terminating equation (1) immediately after refraction at the p th surface,

$$\begin{pmatrix} \xi_p & -x_p \\ \eta_p & -y_p \end{pmatrix} = \begin{pmatrix} \xi_0 & -x_0 \\ \eta_0 & -y_0 \end{pmatrix} \begin{pmatrix} B_{0,p} & D_{0,p} \\ A_{0,p} & C_{0,p} \end{pmatrix}. \dots\dots (3)$$

The accuracy of the values taken for ζ and x is judged by verifying that the conditions

$$\mu^2 = \xi^2 + \eta^2 + \zeta^2$$

and

$$x^2 + y^2 = 2 \int_0^x s dz$$

are satisfied. Assuming that they are not satisfied within the required limits of accuracy, a recalculation is made, using better values of ζ and x . In terms of the old values, the new values ζ' and x' are

$$\left. \begin{aligned} \zeta' &= \frac{1}{2} \left(\zeta + \frac{\mu^2 - \xi^2 - \eta^2}{\zeta} \right) \\ x' &= \frac{x^2 + y^2 + 2 \int x ds}{2s} \end{aligned} \right\} \dots\dots (4)$$

and

As an illustration of the latter, if the surface is a sphere of radius r , the equation is $s+z=r$, and the improved value of z is

$$\frac{x^2+y^2-z^2}{2s} \quad \text{or} \quad \frac{x^2+y^2-z^2}{2(r-z)}.$$

If the surface is a conicoid of revolution of eccentricity e , so that the equation is $s+\epsilon z=r$, where $\epsilon=1-e^2$, the improved value of z is

$$\frac{x^2+y^2-\epsilon z^2}{2s}$$

It will be seen that the new process is a method of successive approximation, and its success depends on its known rapid convergence. The length of the work depends on the closeness of the initially assumed values of z and ζ to the final values. In completely automatic cyclic operations rapidity in reaching the final result is not necessarily a very important or even a desirable requirement. For example, if the initial values are $z=0$, $\zeta=\mu$, i.e. the paraxial values, and each cycle includes the complete instrument, the first round gives the paraxial properties, the second the first-order aberrations, and thereafter groups of aberrations of increasing size are taken into account until higher-order aberrations produce no measurable change. The calculation of a single ray by the new method then gives information that would only be secured on present practice by tracing several rays. On the other hand there are occasions on which only the final result is needed, as when the effect of a change of glass constants or other minor modification of a known system is to be found. In such cases the known values for the previous construction provide suitable initial values for z and ζ , and the modified values may well repeat after a single cycle has been performed. In some circumstances a graphical trace, such as Dowell's (Dowell, 1926), may be a convenient means of deriving approximate initial values for accurate calculations, and with experienced designers "guessing" may be very effective. Several variations of the process may be arranged to suit a variety of conditions.

Like other iterative methods of calculation, this scheme has the property that mistakes are not important, provided they are not incorrect values of fundamental quantities like μ^2 or r . Errors are automatically eliminated—at the worst they lengthen the time taken to reach the final result. The scheme is therefore valuable as a means of checking results reached by other ways of tracing rays. From its very nature it is not less accurate than any other way of tracing rays in which the same number of figures is retained. If, for example, it were required to check a calculation made by the normal N.P.L. method for rays in an axial plane, or to increase the number of significant figures, A would be taken as $(\mu' \cos \phi' - \mu \cos \phi)R$ and z as

$$\frac{2h(\sin \phi - \sin \psi)}{(\cos \phi + \cos \psi)^2 + (\sin \phi - \sin \psi)^2}.$$

(Checking could be made after a single cycle, since

$$\frac{2h(\cos \phi + \cos \psi)}{(\cos \phi + \cos \psi)^2 + (\sin \phi - \sin \psi)^2}$$

is the value of the transverse coordinate of the point of refraction.)

It is worth noting that the one calculation gives the sagittal foci as well as the exact position of the ray. If the sagittal conditions are well satisfied over the field there is no need to consider the tangential conditions, which are merely differentials of the sagittal expressions. If desired, the independent tangential calculation may be made after evaluating $s\xi - x\xi - y\eta$.

Equations (1), (2), (3) and (4) are well suited for the development of formulae for aberrations of different orders.

When this method of tracing rays was first evolved it was intended to apply it to several systems of many types before publishing any description. Owing to war requirements there have not been suitable opportunities for such extensive trials, and the experience so far gained has been limited. In view, however, of the interest now being taken in possible uses of non-spherical refracting and reflecting surfaces, and the importance of developing fully-automatic means of tracing rays, it is thought appropriate to publish this note without further delay.

ACKNOWLEDGMENT

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

APPENDIX

To illustrate the rapidity of convergence of a calculation by this method a marginal ray has been traced with a paraxial ray as the starting point. The example has been taken from Conrady's *Applied Optics* (p 50) as it is likely to be familiar to many readers. The specification of the system and the successive stages from the paraxial ray to the final marginal values are given in table 1. The values of the longitudinal aberration and the change in the focal length from the paraxial value derived at each stage are also given. The values for the former indicate under-corrected aberration at first, changing finally to slight over-correction. The focal-length changes show that the system is not closely corrected for coma.

Table 1

$r_1 = 3.55 = -r_2$, $r_3 = -60$, $t_1 = 0.3$, $t_2 = 0.2$, $\mu_1 = 1.5166$, $\mu_2 = 1.6256$
Successive stages in tracing ray $\zeta_0 = \eta_0 = 0$, $\zeta_0 = 1$, $x_1 = 1$, $y_1 = 0$

Z_1	Z_2	Z_3	ζ_1	ζ_2
0	0	0	1.516600	1.625600
0.140845	-0.132853	-0.007632	1.509618	1.621483
0.143754	-0.142977	-0.007906	1.509215	1.621396
0.143756	-0.143370	-0.007916	1.509214	1.621412
—	-0.143382	—	—	1.621413
				0.991925
Longitudinal aberration			Focal length change	
-0.016819			-0.086012	
-0.000926			-0.072701	
+0.000004			-0.072058	
+0.000039			-0.072036	
+0.000040			-0.072035	

Table 2. Outstanding errors with different methods of calculation

Method	Longitudinal aberration error	Focal length error
5-figure logarithms	-0 000523 (-0 000223)	+0 000047
6-figure polar perpendiculars	+0-000083	+0-000066
8-figure central perpendiculars, no interpolation	+0-000878	+0-000479
8-figure central perpendiculars, full interpolation	+0 000008	+0 000001
7-figure radian tables	+0 000009	+0-000003
6-figure present scheme	0-000000	0 000000

The accuracy attained in calculations according to this scheme is compared with those of other methods in table 2. Unfortunately it was not possible to use Conrady's results to illustrate the accuracy of the normal logarithmic calculation, as his paraxial result has an error of three units in the fourth decimal place. The marginal ray (but not the paraxial ray) was therefore calculated afresh, using Bremmiker's 5-figure tables. Owing to the long last radius the normal formula does not give a very accurate result, and a value derived by using a formula intended to avoid this uncertainty has been added in brackets. The error in the focal length given by this calculation is not typical of that to be expected with 5-figure logarithms: the particular figures of this example happen—quite accidentally—to give a much better result than could be expected.

On the second line are the values obtained with the method normally used at the National Physical Laboratory. The two following lines are those obtained by the original N.P.L. method, using Gifford's sine tables. In the former each angle was read to the nearest second only. The errors involved in this procedure should be about twice those to be expected from the use of 5-figure logarithms, and the abnormally small error in the focal length found in the logarithmic calculation is clearly shown. When interpolation is used, the increased accuracy due to the additional figures is apparent, though it is to some extent diminished through the use of centres of curvature as reference points instead of the surface poles. The radian tables mentioned are some prepared recently at the N.P.L. for use with the B.S.I.R.A. computing scheme. The last line refers to the method described in this paper. All the errors have been determined by comparison with calculations of ten-figure accuracy.

It has been noted earlier that, after the paraxial trace, the changes made in the first cycle represent approximately the first-order aberrations, the second cycle the second and third orders, and so on. The extent to which this holds may be seen from table 3 and the analytical expressions which follow. For this purpose

Table 3. Successive approximations to A , C , ζ_3 , Z_3

A	C	ζ_3	Z_3
0-12567 88492	0-95697 95005	0 99210 24134	-0-00763 17480
0-12705 22669	0-97396 20198	0-99189 60450	-0-00790 55370
0-12683 77581	0-97458 72430	0 99192 34768	-0 00791 56913
0 12682 74221	0-97460 11302	0-99192 47981	-0-00791 59169
0-12682 70639	0-97460 15419	0-99192 48439	-0-00791 59236

values have been calculated to ten decimal places and have been carried far enough to give totals accurate to at least the first six places.

Analytical expressions

$$A = 0.12567\ 88492 + 0.00135\ 95694\ x^2 - 0.00016\ 67749\ x^4 - 0.00003\ 74970\ x^6 \\ - 0.00000\ 60264\ x^8 - 0.00000\ 09097\ x^{10} - 0.00000\ 01357\ x^{12}$$

$$C = 0.95697\ 95005 + 0.01628\ 90110\ x^2 + 0.00119\ 72985\ x^4 + 0.00011\ 90666\ x^6 \\ + 0.00001\ 44030\ x^8 + 0.00000\ 19475\ x^{10} + 0.00000\ 02790\ x^{12}$$

$$\zeta_3 = 1.00000\ 00000 - 0.00789\ 75866\ x^2 - 0.00020\ 20551\ x^4 + 0.00001\ 84401\ x^6 \\ + 0.00000\ 50645\ x^8 + 0.00000\ 08382\ x^{10} + 0.00000\ 01237\ x^{12}$$

$$Z_3 = 0.00000\ 00000 - 0.00763\ 17480\ x^2 - 0.00026\ 02895\ x^4 - 0.00002\ 13407\ x^6 \\ - 0.00000\ 22274\ x^8 - 0.00000\ 02744\ x^{10} - 0.00000\ 00374\ x^{12}$$

Height of intersection with paraxial image plane :—

$$-0.00258\ 03856\ x^3 + 0.00208\ 17289\ x^5 + 0.00042\ 36428\ x^7 \\ + 0.00006\ 80900\ x^9 + 0.00001\ 02632\ x^{11} + 0.00000\ 15177\ x^{13}$$

Path length difference to intersection of ray with paraxial image plane .—

$$0.00024\ 32249\ x^4 - 0.00021\ 62703\ x^6 - 0.00003\ 88728\ x^8 \\ - 0.00000\ 79604\ x^{10} - 0.00000\ 11819\ x^{12}$$

Correction for path differences to paraxial principal focus :—

$$-0.00032\ 42999\ x^3 + 0.00025\ 81211\ x^5 + 0.00005\ 65035\ x^7 \\ + 0.00000\ 88830\ x^9 + 0.00000\ 12493\ x^{11} + 0.00000\ 01673\ x^{13}$$

This example is of interest for another reason. In discussing results obtained by tracing several rays through this lens, Conrady, speaking of analytical methods of calculating higher-order aberrations, makes use of the phrases "... if we went through the tremendous labour of calculating the true secondary aberration by a direct analytical equation ..." and "... the very sensible tertiary aberration (which would utterly defy analytical determination) ...". These views are held by other optical workers. For instance Prof. von Rohr told me that I should only waste my time if I attempted such calculations, because the subject was far too complex for treatment in this way. Such opinions obviously lack scientific justification, and that they are wrong is indicated by the results just given. Admittedly the calculations are tedious, but it is certainly possible to calculate any aberrational coefficients that may be required.

It should be realized that the tables of this appendix are intended solely to indicate the comparative accuracy of the particular mathematical processes. No account has been taken of the accuracy that is necessary or sufficient in these optical computations.

REFERENCES

- DOWELL, J. H., 1926. "Graphical methods applied to the design of optical systems", *Proc. Opt. Conv.* p. 965
SMITH, T., 1915. "On tracing rays through an optical system", *Proc. Phys. Soc.* 27, 502 ; 1918. *Ibid.* 30, 221 ; 1920. *Ibid.* 32, 252 ; 1921. *Ibid.* 33, 174 ; 1930. "The general form of the Smith-Helmholtz equation", *Trans. Opt. Soc.* 31, 241

DIRECTIONAL LOCI IN A MAGNETIC FIELD, AND THE LOCATING OF NEUTRAL POINTS

By DAVID OWEN,

London

MS received 26 February 1945

ABSTRACT A method of exploration of a magnetic field by means of directional loci is described. The points of intersection of any two loci are neutral points. The method is exemplified in the simpler cases of the combination of the earth's field with the field of a bar magnet and with the field of a circular current.

§1. INTRODUCTION

MAGNETIC fields are usually studied by plotting the lines of force in a horizontal plane by means of a compass needle. The procedure for locating a neutral point is usually to track it down by enclosing it within a gradually reduced region bounded by lines of force. The method is ineffective because it becomes impossible to plot the lines accurately close to a neutral point. There is a sort of principle of indeterminacy at play—the nearer the approach to the neutral point, the less accurate the plotting.*

The method here described is to obtain the neutral points as the points of intersection of two continuous curves or loci. The kind of locus considered is that containing all points at which the direction of the field is the same. Any two directions of field may be chosen, but obviously it is advisable to choose directions at right angles to one another; and in cases where the earth's field is involved, these directions should obviously be respectively parallel to and perpendicular to the earth's field. Loci so drawn will be described as the north-south (N-S) locus and the east-west (E-W) locus. Such directional loci are unique. The experimental errors of plotting are non-cumulative, so that the loci can be determined quite definitely. As a point of intersection is approached, the component of field in the chosen direction diminishes to zero and, beyond it, changes sign. The points of intersection are thus points of zero field, or neutral points.

The number of neutral points in any field is even. When the field of a simple bar magnet is superposed on the earth's field, the number of neutral points is two. When the superposed field is that of a circular current, the number of neutral points is either zero or four, according as the strength of the current is below or above a certain critical value. With more complicated fields the number may obviously be greater.

The apparatus required to plot a directional locus consists of a sheet of squared paper on a drawing board and an ordinary small compass needle. Suppose the

* See, however, *Note added in proof* (p. 301).

field to be explored is that of a bar magnet in the earth's field. The squared paper is adjusted on the board so that one set of lines runs magnetically north and south, and the other set therefore east and west. The magnet is now brought up and placed near the centre of the paper with its axis in any desired direction. On moving the compass needle over the paper, some point is soon found where the needle points parallel to the N-S lines. The sharpened point of a pencil is held firmly just above the centre of the needle, the needle is removed, the pencil point moved straight down on to the paper and a dot made. A second point on the locus is chosen about a centimetre away, and so on. Continuing thus, the complete N-S locus is plotted and a continuous graph can be drawn. Similarly the E-W locus can be found. The complete loci will usually extend into all four quadrants round the centre of the magnet. If the experiment is to be restricted to a quick determination of one neutral point, of course only a limited portion of each locus in the neighbourhood of the neutral point is needed. To check the accuracy of setting of the needle at any point of a locus it should be remembered that a small displacement of the needle to one side of the locus will cause an appreciable deviation of direction of one sign, whilst a similar displacement on the opposite side will cause a deviation of opposite sign. If practised at each setting this procedure will ensure correct plotting. With a little experience the loci can be rapidly obtained. The neutral points are thus located with pin-point accuracy with a magnet of reasonable moment, and their distances from the centre of the magnet can be measured within a half per cent.

To illustrate the use of directional loci a few special cases will be considered: first when the field is that of a bar magnet, and then when the field is that of a circular electric current.

§2. BAR MAGNET IN THE EARTH'S FIELD

Since the forms of the loci are easily calculable in the ideal case of a magnetic doublet set with its axis either along or perpendicular to the earth's field, figures showing the forms of these loci will first be given. They will be accompanied by figures showing the loci experimentally plotted in the case of an actual bar magnet. Points on a N-S locus are indicated by a cross (\times), and points on an E-W locus by a centred circle (\odot). The neutral points are at N_1 and N_2 .

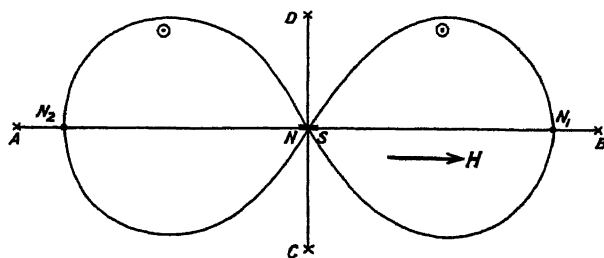


Figure 1. N-S (marked \times) and E-W (marked \odot) loci of magnetic doublet along the earth's horizontal field, S pole to the north, neutral points at N_1 and N_2 .

Case 1. *Axis along the meridian, S pole towards the north.*—The loci for a doublet are shown in figure 1. The N-S locus consists of a pair of straight lines AB and CD along and perpendicular to the axis of the doublet. The E-W locus

is a figure of eight. Denoting the magnetic moment by M and the earth's horizontal field by H , the condition that the field at any point $P(r, \theta)$ shall be in the E-W direction is

$$H - \frac{2M}{r^3} \cos^2 \theta + \frac{M}{r^3} \sin^2 \theta = 0, \quad \text{or} \quad r^3 = M(3 \cos^2 \theta - 1)/H,$$

which is, therefore, the equation of the locus. Putting $\theta = 0^\circ$ or 180° the distance from the doublet to either neutral point is $\sqrt[3]{2M/H}$. At the origin the loops make angles $\cos^{-1} \left(\pm \frac{1}{\sqrt{3}} \right)$ with the axis.

When an actual magnet is used, the N-S locus remains a pair of mutually perpendicular straight lines, along and perpendicular to the axis. Figure 1a

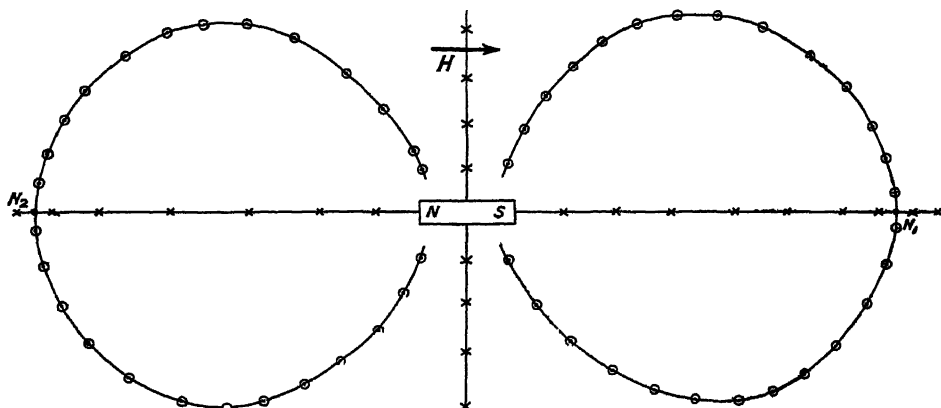


Figure 1a. N-S and E-W loci, and neutral points N_1 , N_2 of 1-in. cobalt-steel magnet along earth's horizontal field, S pole towards the north.

shows also (to half-scale) the two separated branches of the E-W locus. A cylindrical cobalt-steel magnet 1 in. long and $\frac{1}{4}$ in. in diameter was used. The measured distances from the centre of the magnet to N_1 and N_2 are 4.42 in. or 11.22 cm. and 4.40 in. or 11.18 cm., mean = 11.2 cm. (± 5 mm.). Assuming $M = Hr^3/2$, and $H = 0.18$, $M = 126.4$ c.g.s. Assuming $M = \frac{Hr^3}{2} \left(1 - \frac{2l^2}{r^2} \right)$, and taking $l = 0.8 \times$ half-length of magnet, $M = 124.4$ c.g.s.

Case 2. *Axis along the meridian, N, pole to the north.*—Figure 2 represents the case of a doublet. The N-S locus remains as in Case 1. The E-W locus consists of two loops, the long axis now running east and west, so that N_1 and N_2 lie on the east-west line. The equation of the E-W locus is

$$\frac{M}{r^3} (1 - 3 \cos^2 \theta) = H.$$

The lengths ON_1 and ON_2 to the centre of the magnet are equal to $\sqrt[3]{\frac{M}{H}}$. At the origin the loops cross the axis at angles $\cos^{-1} \left(\pm \frac{1}{\sqrt{3}} \right)$.

Using the same cobalt-steel magnet as before, the observed plot is shown (to half-scale) in figure 2a. From this $ON_1 = 3.48$ in. = 8.84 cm., $ON_2 = 3.46$ in. = 8.79 cm.,

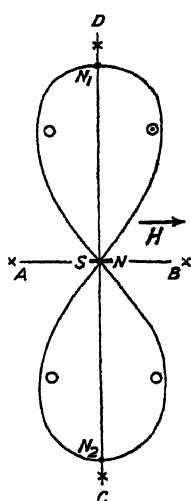


Figure 2 N-S (straight lines AB and CD) and E-W (figure of eight) loci of magnetic doublet along earth's horizontal field, N pole to the north Neutral points at N_1 and N_2 .

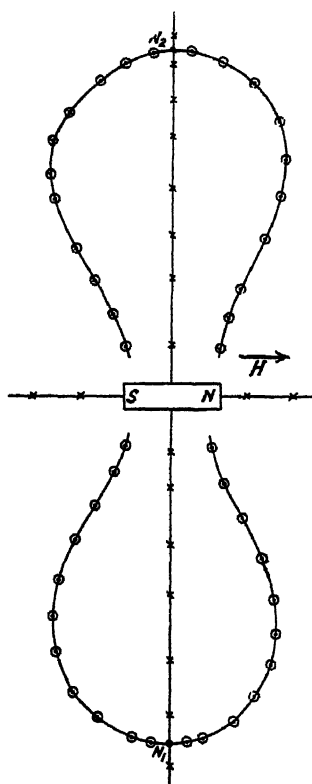


Figure 2 a. N-S and E-W loci of 1-in. cobalt-steel magnet, N pole to the north, neutral points at N_1 , N_2 .

mean value = 8.81 cm. Assuming the formula $M = Hr^3$, the value of M is 123.0 c.g.s., and assuming $M = Hr^3 \left(1 + \frac{3}{2} \frac{l^2}{r^2}\right)$ the value of M is 125.3 c.g.s.

Case 3. *Axis perpendicular to the meridian.*—The case for the doublet, with N pole to the east, is shown in figure 3. The N-S locus is the pair of straight

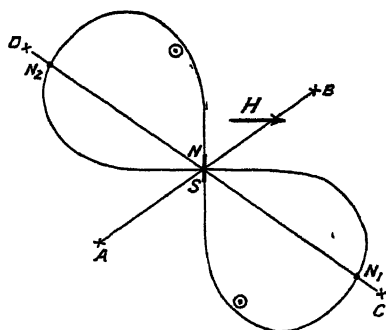


Figure 3. N-S locus (straight lines AB and CD) and E-W locus (figure of eight) of magnetic doublet with axis \perp to earth's field, N pole to the west, neutral points at N_1 and N_2 .

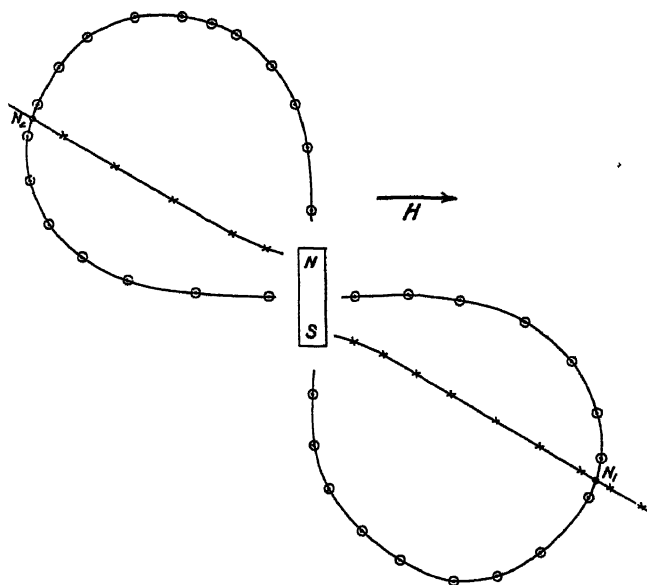


Figure 3a. N-S and E-W loci and neutral points N_1 and N_2 of 1-in cobalt-steel magnet, axis \perp to H , N pole to the west

lines AB and CD making angles $\tan^{-1} \sqrt{2}$ (nearly 55°) with the axis. The equation of the E-W locus is

$$r^3 = \frac{3}{2} \frac{M}{H} \sin 2\theta.$$

It is a figure of eight whose axis makes an angle of 45° with the magnetic axis, and it touches the axis of the doublet and the equatorial line.

The experimental graphs for the same 1-in. cobalt-steel magnet are shown in figure 3 *a*. Near the magnet the N-S locus deviates appreciably from the straight line pertaining to the doublet.

§ 3 FIELD OF AN ELECTRIC CURRENT

The ideal case of an infinitely long straight conductor of circular section, placed vertically in the earth's field, and carrying a steady current, allows of a simple and suggestive treatment. In figure 4 the smaller circle of radius $OA = a$

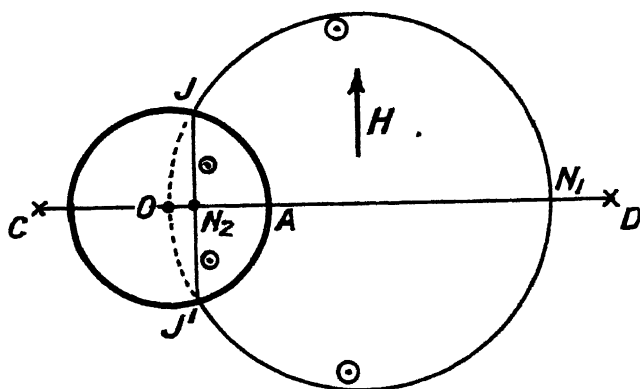


Figure 4. Vertical circular conductor, of radius OA , carrying a downward electric current (exceeding the critical value) in the earth's field. The N-S locus is the straight line COD . The E-W locus is the portion of the larger circle on ON_1 as diameter which is external to the conductor, together with the straight line JJ' inside the conductor. There is an external neutral point at N_1 and an internal neutral point at N_2 .

represents a section of the conductor. No neutral points exist unless the current I exceeds a limiting value $\frac{1}{2}Ha$. With any current exceeding this value, the current being downwards, an external neutral point N_1 occurs on the east side, such that $2I/ON_1 = H$. A second neutral point exists at N_2 within the conductor, such that $\frac{2I \cdot ON_2}{a^2} = H$. Thus $ON_1 \cdot ON_2 = a^2$. The N-S locus is the straight

line running east and west through the centre O of the conductor. For points outside the conductor the E-W locus is the portion of the circle on ON_1 as diameter lying outside the conductor. For points inside the conductor, the E-W locus is the straight line J_1J_2 joining the points where the circle on ON_1 cuts the conductor. At the limiting value of the current the neutral points coincide at A on the surface of the conductor.

The case of a circular coil, or rather a short helical coil with horizontal axis, will next be considered. The determination of either directional locus does not in general lend itself to easy theoretical treatment. There is again obviously a limiting value of current below which no neutral points exist. For currents exceeding this value there are four neutral points, two within the conductor and two external to it.

Case 1. *Axis of the coil parallel to H , field due to the current in the same direction as H at the centre of the coil.*—In this case the N-S locus consists of a pair of straight lines, namely, the axis of the coil and the transverse median line.

The E-W loci obtained experimentally are shown to half-scale in figure 5, for currents of 1, 2 and 3 amp., only the branches on one side of the coil being drawn. The corresponding external neutral points are at N' , N'' and N''' . The mean radius of the coil is 3.76 in. and its axial width 0.5 in.

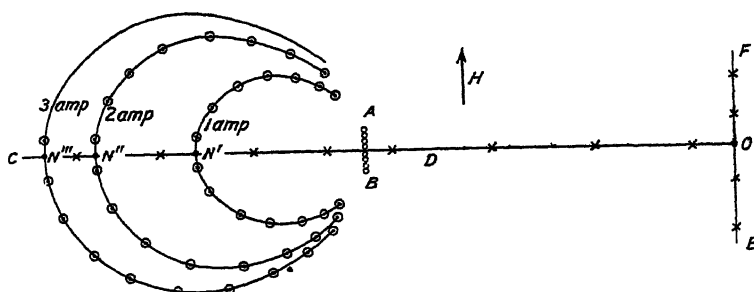


Figure 5 Electric currents through circular coil with axis along the earth's horizontal field, fields on axis in *same* direction as H . AB =section of the coil of 8 turns, centre at O . The N-S locus consists of the pair of straight lines CD and the axis of the coil. The branches of E-W loci are shown for currents of 1, 2 and 3 amp., the neutral points being respectively at N' , N'' and N''' . (Complete figure is symmetrical about O . Right-hand half omitted.)

Case 2. *Axis of the coil parallel to H , field due to the current opposed to H at centre of coil.*—The experimental graphs for this case are seen, to half-scale, in figure 6 for six values of the current as indicated on the curves. The complete

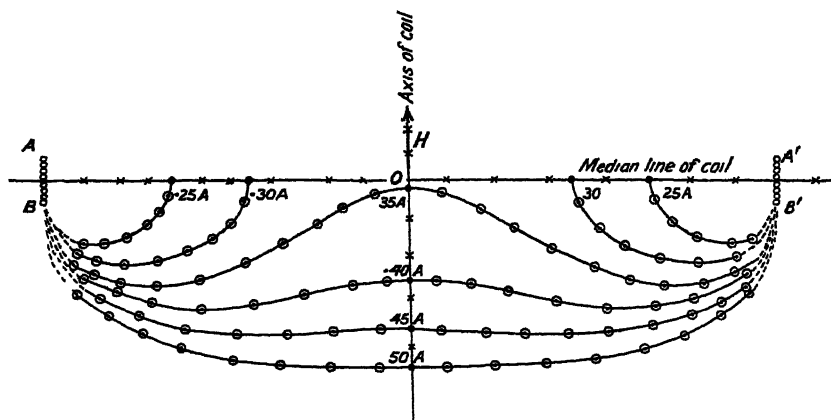


Figure 6 Circular coil of 8 turns, axis along earth's field. Current in direction producing field at centre opposed to earth's field H . N-S loci along axis of coil and its median line. The E-W loci are the curves marked \odot . Neutral points marked \bullet are on axis of coil for currents 0.35 A., 0.40 A., 0.45 A., 0.50 A., and on median line for currents 0.30 A. and 0.25 A. (Neutral points coincide at centre O of coil for calculated current 0.345 A.) Radius of coil 3.76". (Upper half of complete figure, which is symmetrical about median line, is omitted.)

curves would be symmetrical about the median line of the coil, but only the halves on one side of that line are drawn. For currents of 0.35 amp. and upwards two neutral points lie on the axis of the coil, one on either side of the centre. They coincide at the centre for a current of 0.345 amp. For currents below this

last value the two neutral points lie on the median line, on either side of the centre. There are two additional neutral points within the conductor.

Case 3. *Axis of the coil perpendicular to H .*—This case is given as an example of a less symmetrical setting of the coil with respect to the earth's field. Figure 7

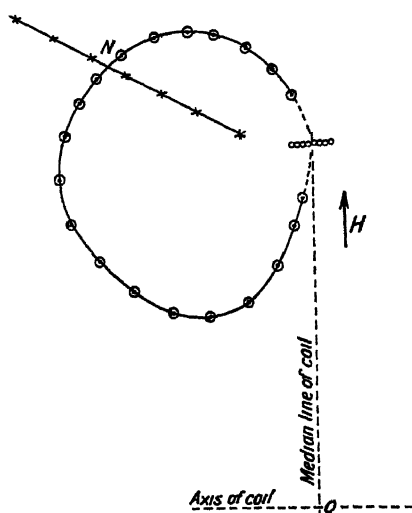


Figure 7 Circular coil of 8 turns, axis perpendicular to earth's field H . Current of 10 amp., producing field from west to east along axis. Only one branch of N-S locus (marked \times) and of E-W locus (marked \circ) shown. Neutral point at N (the second neutral point should be on line joining N to centre O of coil, at equal distance on other side of centre.)

shows (to half-scale) one branch only of each of the loci, thus determining one of the two external neutral points. The value of the current was 1 amp., and the coil was that used for the previous cases.

Note added in proof

Since this paper was presented, it has been noticed that there are two special lines of force which provide exceptions to the general rule enunciated in §1. This pair of lines pass through the neutral point itself. They will be the subject of a brief paper to follow.

A DIFFERENTIAL METHOD OF ADJUSTING THE ABERRATION OF A LENS SYSTEM

BY A. L. M'AULAY AND F. D. CRUICKSHANK,
Tasmania

MS. received 25 January 1944

ABSTRACT A method is described for computing a highly corrected camera lens system when the general type is given, and a rough design has been made

The method separates into two parts. First, changes in the rays incident at any surface are expressed as functions of alterations of the surface constants, secondly, the ray changes are transferred to the final image space and expressed as changes in the aberrations of the system. In this way a table is obtained by the use of which desired modifications of the aberrations can be achieved by suitable surface alterations.

§ 1. INTRODUCTION

THE need for the manufacture of camera lenses of a wide range of characteristics became urgent in Australia in 1942. It was necessary to use the stocks of optical glass available in the country. Local requirements and facilities precluded the possibility of producing anything approaching a direct copy of overseas lenses, so that the general problem of design had to be faced. Preliminary work showed that an attempt to obtain the final design of a highly corrected lens system by successive ray tracings was unsystematic and slow. Such methods could not be tolerated in an emergency, and hence a new approach had to be found. It is believed that a considerable advance towards a faster and more systematic method of completing a rough design has been made, and a general outline of it is presented in this paper

The method described has made possible the design and construction of highly corrected camera lenses from rough data which gave the type only. The work has been developed rapidly by a group with no previous experience in lens design. It is this technical success that gives the method merit. The method is a logical addition to the standard trigonometrical ray-trace, and serves to raise the whole status of the trigonometrical method to that of a powerful and systematic tool for use in the final stages of an optical design. It has been said that the trigonometrical method is empirical and uninformative, but it will be shown that a most instructive analysis of any optical system can be made on the basis of a ray trace, and that, in particular, a complete and exact description of all the general tendencies of the system is readily available at the price of some computations which can be undertaken by computers without a knowledge of technical optics.

§ 2 THE BASIC PRINCIPLES OF THE METHOD

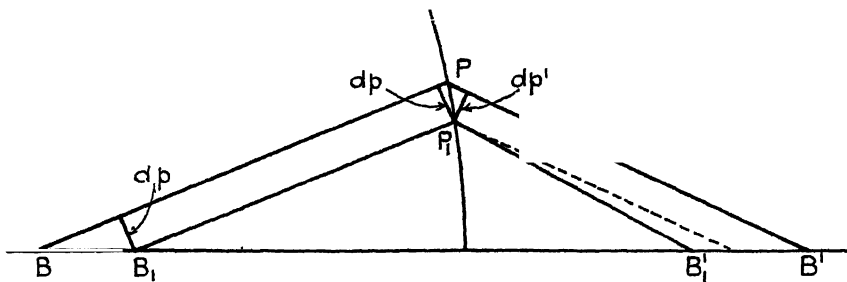
The primary object of a ray trace is to obtain an analysis of the aberrations of the system. The aberration measures are defined in terms of certain intersection points associated with selected traced rays in the final image space, as described,

for example, by Conrady (1929). If the aberrations revealed by the trace of a rough design are too large, the designer's problem is to diminish them by appropriate changes in curvature, thickness, and refractive index within the system.

The ray trace provides a complete specification of the paths followed by the selected rays through the system. Suppose now that a change is made at surface i of the system. As a result, each of the selected rays will be refracted at this surface slightly differently, and their paths through the remainder of the system will differ from the original traced paths. Each ray will finally leave the last surface of the system at a point displaced a small distance from the emergence point of the corresponding traced ray, and in a direction inclined at a small angle to that of the traced ray. The intersection points associated with the emergent rays, and hence the aberrations of the system, will now be changed. The essence of the present method is the development of an analysis which will permit the computation of the aberration changes due to a small change made at any surface of the system. In doing this, the detailed information from the trace is used, and the traced path of each ray is used as a convenient reference frame relative to which is described the new path of the ray resulting from the particular change. For this purpose, certain auxiliary variables are chosen, and by means of changes in these, the new path of the ray can be described relative to the original path. For the variables of the trace generally we use the notation of Conrady (1929). Several sets of variables may be selected from among the quantities $L, L', U, U', I, I', \dots$ of the trace, and the relative merits of each set could be debated. The writers have found after some experience in the use of the method that the differentials dU, dU', dp, dp' , as defined in the next section, are simple to use and lead to expressions which have certain advantages in computation.

§3 SPECIFICATION OF THE CHANGED RAY PATH AT A SINGLE SPHERICAL REFRACTING SURFACE

Suppose that the path of a selected incident ray is traced trigonometrically through a single spherical refracting surface. The ray path is specified by the quantities (L, U) and (L', U') and is represented by $B P B'$ in the figure.



$B P B'$ is the traced ray path Line through new incidence point parallel to traced refracted ray shown broken $\sin U$ and dp are negative, as shown

Change in incidence point. Suppose that the incidence point of the ray is changed from P to P_1 , the direction of the ray remaining unchanged. The path

of the ray through the surface would now be represented by $B_1P_1B_1'$. We propose to specify the change in incidence point by the perpendicular distance, dp , between the two rays, the quantity being positive when the new parallel ray lies above the original. The new ray will have an intersection length $L + dL$, compared with L for the original path, and

$$dp = \sin U \cdot dL. \quad \dots\dots(1)$$

Notice that dp is not the total differential of the familiar quantity $L \sin U$. We are never particularly interested in the integral quantity p , except possibly in a formal way, but it is important to stress that relative to a given ray path, (L, U) , which intersects the refracting surface, the quantity $dp = \sin U \cdot dL$ uniquely defines a neighbouring incidence point.

On the other side of the surface it is natural to specify the same incidence-point displacement by another quantity, dp' , which is the perpendicular distance between the original refracted ray and a line parallel to this drawn through the new incidence point. The relation between dp and dp' for small displacements of the incidence point is

$$dp' = (\cos I' / \cos I) \cdot dp$$

$$\text{or} \quad \frac{\partial p'}{\partial p} = \frac{\cos I'}{\cos I}. \quad \dots\dots(2)$$

Change of direction. A change in the direction of the ray incident at P is specified by dU , and the resulting change in the direction of the refracted ray by dU' . It will now be seen that any new path differing by small amounts in direction and in location of incidence point from the original ray-path can be specified relative to the original path by the quantities dp , dU , on the incidence side of the surface, and by the corresponding quantities dp' , dU' , after refraction at the surface.

§ 4. THE EFFECT OF SMALL CHANGES MADE AT A SINGLE SPHERICAL REFRACTING SURFACE

We consider now the effect on the direction of the refracted ray of various changes which can be made at a single surface at which an incident ray is refracted. Writing down the standard ray-tracing equations in the form

$$(L - r) \sin U = r \sin I, \quad \dots\dots(3)$$

$$N' \sin I' = N \sin I, \quad \dots\dots(4)$$

$$U + I = U' + I', \quad \dots\dots(5)$$

we differentiate (4) and (5), putting $N/N' = n$, and obtain

$$dU' = dU - \left(n \frac{\cos I}{\cos I'} - 1 \right) dI - \frac{\sin I}{\cos I'} dn. \quad \dots\dots(6)$$

If the direction of the incident ray changes from U to $U + dU$, the change in the angle of incidence, I , is $dI = -dU$, and hence, from (6),

$$dU' = \frac{\partial U'}{\partial U} dU = dU - \left(n \frac{\cos I}{\cos I'} - 1 \right) (-dU),$$

$$\text{whence} \quad \frac{\partial U'}{\partial U} = n \frac{\cos I}{\cos I'}. \quad \dots\dots(7)$$

If the position of the incidence point is changed by an amount dp , the change in I , obtained by differentiating (3) with U and r constant, is

$$dI = \frac{\sin U \cdot dL}{r \cos I} = \frac{dp}{r \cos I},$$

which, on combination with (6), gives

$$\frac{\partial U'}{\partial p} = \left(1 - \frac{\partial U'}{\partial U}\right) / (r \cos I). \quad \dots\dots(8)$$

For a small variation in the relative refractive index, n , equation (6) gives at once

$$\frac{\partial U'}{\partial n} = -\frac{\sin I}{\cos I'}. \quad \dots\dots(9)$$

If a small change is made in the curvature, c , of the surface, the change in I , obtained by differentiating (3) with U constant, is

$$dI = \frac{L \sin U}{\cos I} \cdot dc,$$

which, on combination with (6), gives

$$\frac{\partial U'}{\partial c} = \left(1 - \frac{\partial U'}{\partial U}\right) \frac{L \sin U}{\cos I}. \quad \dots\dots(10)$$

§ 5 THE TRANSFER COEFFICIENTS

We consider next the problem of the effect on the complete ray-path of a small change made at one surface of a system of k spherical refracting surfaces. Suppose that the path of a selected incident ray has been traced through the system. After refraction at any surface i of the system, the ray-path is specified by the quantities L_i' , U_i' , in the usual way. A small change is now made at this surface such that the ray, after refraction at the same incidence point on the surface, takes a new direction specified by $U_i' + dU_i'$. The path of the ray through the remainder of the system will be different from the original traced path, and finally the ray will emerge from the last surface at a point displaced by some amount, dp_k' , from the emergence point of the traced ray, and in a direction inclined at an angle, dU_k' , to that of the traced ray. We may express dU_k' and dp_k' in terms of dU_i' by writing

$$dU_k' = \frac{\partial U_k'}{\partial U_i'} \cdot dU_i', \quad \dots\dots(11)$$

$$dp_k' = \frac{\partial p_k'}{\partial U_i'} \cdot dU_i'. \quad \dots\dots(12)$$

In a similar manner, if a small change is made at surface i resulting only in an incidence-point change, dp_i , the inclination angle, U_i , of the ray remaining unchanged, the ray will emerge from the last surface at a point and in a direction specified by some other values dp_k' , dU_k' , relative to the emergent traced ray. In this case we may write

$$dU_k' = \frac{\partial U_k'}{\partial p_i} \cdot dp_i, \quad \dots\dots(13)$$

$$dp_k' = \frac{\partial p_k'}{\partial p_i} \cdot dp_i. \quad \dots\dots(14)$$

The equations (11) to (14) define the *transfer coefficients* $\partial U'_k/\partial U'_i$, $\partial p'_k/\partial U'_i$, $\partial U'_k/\partial p_i$, $\partial p'_k/\partial p_i$, which can be calculated for a traced ray at each surface of the system. They determine the effect of a dU' -change or a dp -change made at any surface on the path of the ray as it leaves the last surface of the system. The transfer coefficients for dU' -changes and for dp -changes are the fundamental transfer quantities from which all others may be conveniently derived. It may be thought confusing to use the pair of differentials dU' and dp for the description of small changes in the system, one of these relating to a quantity before refraction and the other to a quantity after refraction of the ray at the surface, but in practice any change which is made in a system can be described in terms of either a single dU' -change or a single dp -change, so that the transfer coefficients selected are not only sufficient for the present problem, but also require less computation than any other set. It is to be stressed that what is in view is to fashion a powerful designing tool, considerations of mathematical elegance being sacrificed to this end.

It will be shown in a subsequent paper that certain relations connect the transfer coefficients at two successive surfaces of the system, such that if the values of the coefficients are known for surface $(i+1)$ the coefficients for surface i may be calculated from them. At the last surface of the system it is obvious that the transfer coefficients have simple values, for

$$\partial U'_k/\partial U'_k = 1; \quad \partial p'_k/\partial U'_k = 0,$$

while $\partial p'_k/\partial p_k$ and $\partial U'_k/\partial p_k$ are calculated from equations (2) and (8) respectively. The computation of the transfer coefficients can thus commence from the last surface and continue backwards, surface by surface, through the system until the first surface is reached.

§ 6 THE TRANSFER COEFFICIENTS FOR THE INTERSECTION POINTS

The next step in the development of the method is to determine the changes in the positions of the intersection points in the final image-space associated with the traced rays due to the differential changes made within the system. In a normal trace it is usual to select three rays of the axial pencil, the marginal, zonal and paraxial rays. Associated with these rays are the intersection points $M(L'_m, 0)$, $Z(L'_z, 0)$, $P_x(L'_x, 0)$ of each of these rays with the principal axis, their positions being specified by rectangular coordinates relative to the pole of the last surface. Omitting the detail, given in a subsequent paper, it can be shown that

$$\frac{\partial L'_m}{\partial U'_i} = \left(\frac{\partial p'_k}{\partial U'_i} - S'_k \cdot \frac{\partial U'_k}{\partial U'_i} \right)_{mr} \operatorname{cosec} U'_{mk} = (C_{U'_i})_m \operatorname{cosec} U'_{mk}, \dots\dots (15)$$

$$\frac{\partial L'_m}{\partial p_i} = \left(\frac{\partial p'_k}{\partial p_i} - S'_k \cdot \frac{\partial U'_k}{\partial p_i} \right)_m \operatorname{cosec} U'_{mk} = (C_{p_i})_m \operatorname{cosec} U'_{mk}, \dots\dots (16)$$

in which S'_k is the distance measured along the ray from the last surface to the intersection point, and the C -quantities are introduced as convenient abbreviations for the bracketed functions of the transfer coefficients in the first expressions. Corresponding expressions hold for the intersection points Z and P_x .

For an oblique pencil it is usual to trace a set of three rays denoted by the symbols a , pr , and b in the Conrady notation. Associated with these rays are the points $Ab(L'_{ab}, H'_{ab})$, the intersection of the rays a and b ; $Pr(L'_{pr}, H'_{pr})$, the point at which the principal ray cuts the plane through Ab at right angles to

the axis; $\text{Prf}(l', H'_{prf})$, the point at which the principal ray strikes the plane through the paraxial focus at right angles to the axis. Finally, though not related to a traced ray, there is the ideal image point, $\text{Id}(l', H'_{id})$. This point shows where the principal ray would strike the paraxial image plane if the paraxial value of the magnification held throughout the field.

For the point Ab it can be shown that

$$\frac{\partial L'_{ab}}{\partial U'_a} = (C_{U'_i})_a \operatorname{cosec}(U'_a - U'_b)_k \cos U'_{bk}, \quad \dots\dots(17)$$

$$\frac{\partial H'_{ab}}{\partial U'_a} = -(C_{U'_i})_a \operatorname{cosec}(U'_a - U'_b)_k \sin U'_{bk}, \quad \dots\dots(18)$$

with corresponding expressions for changes in the b ray.

For the points Pr, Prf, it follows similarly that

$$\frac{\partial H'_{pr}}{\partial U'_i} = (C_{U'_i})_{pr} \sec U'_{prk}, \quad \dots\dots(19)$$

$$\frac{\partial H'_{prf}}{\partial U'_i} = (C_{U'_i})_{prf} \sec U'_{prk}. \quad \dots\dots(20)$$

Exactly similar relations hold for the transfer coefficients for small dp -changes.

§7. THE TRANSFER COEFFICIENTS FOR ACTUAL SURFACE CHANGES

In an actual design, the variables at the disposal of the designer are surface curvature, refractive index, and the axial separation between successive surfaces. For our purposes the axial separation is best specified by the symbol d . In order to change the separation, d_i , between the surfaces $(i-1)$ and i by an amount ∂d_i , the surface i and all succeeding surfaces are moved through ∂d_i . From the equations of the preceding section it follows at once that for the axial points, M, Z, and Px,

$$\frac{\partial L'}{\partial c_i} = C_{ci} \operatorname{cosec} U'_k; \quad \dots\dots(21)$$

for the point Ab,

$$\frac{\partial L'_{ab}}{\partial c_i} = [(C_{ci})_a \cos U'_{bk} - (C_{ci})_b \cos U'_{ak}] \operatorname{cosec}(U'_a - U'_b)_k, \quad \dots\dots(22)$$

$$\frac{\partial H'_{ab}}{\partial c_i} = -[(C_{ci})_a \sin U'_{bk} - (C_{ci})_b \sin U'_{ak}] \operatorname{cosec}(U'_a - U'_b)_k; \quad \dots\dots(23)$$

while for the points Pr, Prf,

$$\frac{\partial H'_{pr}}{\partial c_i} = (C_{ci})_{pr} \sec U'_{prk} - \frac{\partial L'_{ab}}{\partial c_i} \tan U'_{prk}, \quad \dots\dots(24)$$

$$\frac{\partial H'_{prf}}{\partial c_i} = (C_{ci})_{prf} \sec U'_{prk} - \frac{\partial l'}{\partial c_i} \tan U'_{prk}. \quad \dots\dots(25)$$

Corresponding expressions for the transfer coefficients with respect to n and d are obtained by replacing c throughout by n or d . The C -quantities entering into these expressions are defined by

$$C_{ci} = C_{U'_i}(\partial U'_i / \partial c_i), \quad \dots\dots(26)$$

$$C_{ni} = C_{U'_i}(\partial U'_i / \partial n_i), \quad \dots\dots(27)$$

$$C_{di} = C_{p_i}(\partial p_i / \partial d_i). \quad \dots\dots(28)$$

Direct differentiation of the expression for the ideal image-point yields the required expressions for the transfer coefficients for this point. Thus for an infinitely distant object

$$H'_{id} = -f' \tan U_{pr},$$

$$\text{whence} \quad \frac{\partial H'_{id}}{\partial c_i} = \frac{y}{u_k'^2} \tan U_{pr} \frac{\partial u_k'}{\partial c_i} \frac{\partial u_i'}{\partial c_i}, \quad \dots\dots(29)$$

U_{pr} being the inclination angle of the principal ray at the first surface. The differential coefficients involved are those normally computed for the paraxial ray. Similar expressions hold for the transfer coefficients with respect to n and d .

§ 8 THE TRANSFER COEFFICIENTS FOR THE ABERRATIONS

The last stage of the development can now be made by writing down expressions for the rate of change of each aberration with respect to the curvature, refractive index, and axial separation. Thus for the spherical aberration

$$\frac{\partial LA'}{\partial c_i} = \frac{\partial l'}{\partial c_i} - \frac{\partial L'}{\partial c_i}, \quad \dots\dots(30)$$

for the curvature of the tangential field

$$\frac{\partial X_T'}{\partial c_i} = \frac{\partial L'_{ab}}{\partial c_i} - \frac{\partial L'}{\partial c_i}, \quad \dots\dots(31)$$

for the tangential coma

$$\frac{\partial H'_{pr}}{\partial c_i} - \frac{\partial H'_{ab}}{\partial c_i} \quad \dots\dots(32)$$

and for the distortion

$$\frac{\partial dist'}{\partial c_i} = \frac{\partial H'_{id}}{\partial c_i} - \frac{\partial H'_{prf}}{\partial c_i}. \quad \dots\dots(33)$$

If additional zonal and principal rays for the colours r and v , for which the system is achromatized, are traced, the chromatic aberrations can be brought into the general scheme, giving

$$\frac{\partial Lch'}{\partial c_i} = \frac{\partial L'_{pr}}{\partial c_i} - \frac{\partial L'_{rv}}{\partial c_i}, \quad \dots\dots(34)$$

$$\frac{\partial Tch'}{\partial c_i} = \frac{\partial H'_{prr}}{\partial c_i} - \frac{\partial H'_{prv}}{\partial c_i}. \quad \dots\dots(35)$$

Corresponding expressions hold for transfer coefficients with respect to n and d . A computation of these coefficients, which measure the aberration change per unit change in curvature, axial separation, and refractive index at any surface of the system, provides the designer with complete information for the final balancing of the aberrations of the system. It provides a very complete analysis of the tendencies of the system. The expressions given above cover the general aberrations deducible from the tracing of tangential rays. The method is capable of extension in various directions to include aberrations which are of special importance in particular cases.

The transfer coefficients developed for changes of refractive index refer to changes at one surface. In practice, any change of glass involves two surfaces. By compounding the effects of simultaneous changes at two surfaces we obtain a series of transfer coefficients giving the changes in the aberrations, other than

the chromatic aberrations, for a refractive index change for any *component* of the system. The colour aberrations being particularly related to the dispersions of the glasses, we develop transfer coefficients for these aberrations with respect to dispersion changes. The transfer coefficients for glass changes form a most valuable part of the general machinery of the method.

A degree of freedom not so far discussed is the position of the diaphragm. The effect of a differential displacement of the diaphragm is easily described in terms of the transfer coefficients for *d*-changes at the first surface of the system. As an additional piece of information, the transfer coefficients for a *dp*-change of the principal ray at the first surface give the *t*-focus of the system.

§9 THE METHOD IN PRACTICE

As an example of the use of the foregoing method, a detailed account of one particular stage in the correction of a certain photographic lens is given. The example is chosen at random, but gives a fair idea of the order of accuracy of the method when a number of changes are made simultaneously, some of these being not of differential size. The trace of the rough design at the particular stage of development gave the aberrations in the first line of table 1, an axial pencil and two oblique pencils having been traced at 22° and 32° respectively.

Table 1

LA_m'	LA_z'	X'_{T22}	coma'_{T22}	dist'_{22}	X'_{T32}	coma'_{T32}	dist'_{32}
-1.256	-0.0600	-1.185	-0.0014	0.0085	-3.520	-0.1280	0.0150
-0.624	0.210	0.407	-0.0047	-0.0130	0.010	-0.0530	0.0840

The step undertaken at this stage was to attempt to reduce the heavy negative curvature of field at 22° and 32°, to improve the over-corrected spherical aberration, and to hold, if possible, the well-corrected coma and distortion at 22°. A study of the transfer coefficients of the system showed that there were surfaces providing opportunity for a general move in this direction. The surfaces selected had the coefficients given in table 2.

Table 2

Surface	LA_z'	X'_{T22}	coma'_{T22}	dist'_{22}	X'_{T32}	coma'_{T32}	dist'_{32}
1 $\partial/\partial c$	1879.3	1003.1	-19.19	82.68	1299.3	-6.24	284.64
4. $\partial/\partial c$	510.0	-707.9	-110.5	44.30	-1673.5	-181.3	166.96
5. $\partial/\partial d$	-0.0135	0.3417	-0.0035	-0.1013	0.2294	-0.0039	-0.3981
6 $\partial/\partial d$	-0.0075	0.7681	-0.0093	-0.0094	1.808	0.0034	-0.0479
7. $\partial/\partial d$	0.0025	-0.2663	0.0091	0.0772	-0.9908	0.0384	-0.2839

After a careful study of these coefficients, the following changes were made.

$$\begin{aligned} \delta c_1 &= 0.0001653, & \delta d_5 &= -0.613, & \delta d_7 &= -0.817, \\ \delta c_4 &= -0.0001987, & \delta d_6 &= 1.263. \end{aligned}$$

The second line of table 1 gives the residual aberrations found by a trace of the system after alteration. The large amounts of curvature have been removed, rather overshooting the mark at 22° , the coma and distortion at 22° are still inappreciable, while the coma at 32° and the spherical aberration have moved appreciably in the right direction.

Although the coefficients are exact to first order for single changes of differential magnitude only, they give a very good approximation to the effects of a number of larger changes made simultaneously in the system. Naturally, it is to be expected that as the aperture of the system becomes greater the approximation to the aberration changes becomes less reliable, but a series of steps enables the correction to be carried out successfully.

REFERENCE

CONRADY, 1929. *Applied Optics and Optical Design* (Oxford).

IMPERFECTIONS OF CRYSTAL LATTICES AS INVESTIGATED BY THE STUDY OF X-RAY DIFFUSE SCATTERING

By A. J. GUINIER,

Laboratoire d'essais du Conservatoire National des Arts et Métiers, Paris

MS received 7 April 1945

ABSTRACT. A perfect crystal diffracts x rays only in discrete directions, given by the Bragg-Laue laws. All scattering observed in directions other than these, except Compton scattering, is due to imperfections of the crystal. The experimental method of obtaining useful scattering patterns and the principles of the calculation of actual structure from the distribution of anomalous scattering are described. Examples of application of this method to the study of thermal atomic movements, lattice deformations under mechanical stresses, defects of periodicity in diamond and in fibrous crystals, space-arrangement of atoms in solid solutions (order-disorder transformation, age-hardening) and scattering by small particles are briefly outlined.

§1 INTRODUCTORY

THE theory of the diffraction of x rays by a small crystal is based upon the hypothesis that this crystal is strictly periodic, that is to say, any given atom and all its homologues are situated just at the nodes of a perfect geometric lattice. Then calculation shows that each system of the network of planes can reflect the incident rays according to the optical laws, but only for fixed incidence angles given by the well known Bragg law. In other directions, interferences annul the diffracted radiation. Thus a perfect lattice does not give rise to any scattering out of the selective reflexion directions, except for the continuous Compton scattering, which is very weak for the relatively long wavelengths we are considering. We may therefore enunciate the principle that *all diffuse scattering proceeds from some imperfections of the crystal lattice.*

Our purpose here is to show how this scattering is theoretically related to displacements of the atom, how it can be experimentally detected, and also to give a general survey of the various studies carried on in our Laboratory on this subject.

In the following theoretical treatment of the subject, we shall rely largely on a recent paper by Ewald (1940).

Let us consider a diffracting body (figure 1); let $\rho(\vec{x})$ be the electronic density at the extremity M of the vector \vec{x} ; let \vec{S}_0 and \vec{S} be the unit vectors along the

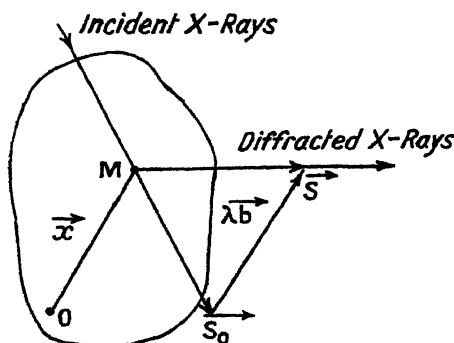


Figure 1 Diffraction of x rays

incident and diffracted rays, and let us call \vec{b} the vector $\frac{\vec{S}-\vec{S}_0}{\lambda}$. If A_e is the amplitude of the wavelet diffracted by one electron, according to the Thomson formula, the amplitude of the wave diffracted by the whole body is

$$F(\vec{b}) = A_e \iiint_{(v)} \rho(\vec{x}) e^{-2\pi i \vec{b} \cdot \vec{x}} dv_x. \quad \dots\dots(1)$$

This equation states that $\frac{F(\vec{b})}{A_e}$ is the Fourier transform of $\rho(\vec{x})$, being defined for all vectors \vec{b} of a space, called the *reciprocal space*. The formula (1) permits the calculation of the diffraction pattern for any given distribution of matter in the diffracting body.

If $\rho(\vec{x})$ is a periodic function, i.e. if the crystal lattice is perfect, $F(\vec{b})$ is zero at every point of the reciprocal space, except at a set of points which are the nodes of the reciprocal lattice, each one corresponding to the selective reflection upon one set of reticular planes. This is another and more general form of Bragg's law. If on the other hand the crystal is perturbed, $\rho(\vec{x})$ is no longer periodic, and $F(\vec{b})$ has a finite value for any vector \vec{b} . In fact, the lattice imperfections we consider are small perturbations of the perfect lattice. Then, as for a perfect crystal, the function $F(\vec{b})$ remains considerable at the nodes of the reciprocal lattice, but now is not zero between these points. This corresponds to the scattering observed in addition to the Bragg-Laue spots, due to the normal selective reflexions.

imperfections is not easy. We cannot deal here with the mathematical calculations involved. We will only show what results can be obtained. But, as an example, we shall give the application to a particular case of a general method of calculation, similar to Patterson's method used in the study of structures.

§ 2. THE EXPERIMENTAL PROCEDURE

What are the conditions necessary for obtaining useful scattering patterns? First of all, the anomalous scattering is generally very weak when compared to the selective reflexion. Moreover, the scattering areas may be very extensive and possess complicated forms. It follows that

1. The method must be such that crystals with perfect lattice do not give any scattering, except the Compton scattering; secondary fluorescence radiation, and scattering originating outside the specimen, must be eliminated. Above all, the incident radiation must be strictly monochromatic, for, if the principal radiation were accompanied by a continuous spectrum, the diffraction effects of the latter would be much greater than the scattering to be studied.

2. In spite of the monochromatization, the incident beam must be very intense, for the phenomena we are looking for are very weak.

3. In most cases the sample must be a single crystal held in a fixed position during the exposure. With polycrystalline samples, the scattering observed in each direction is an average value and the pattern becomes generally very difficult to interpret. To sum up, for a scattering pattern, one must use a *monochromatic beam* and an *immovable crystal*. The conditions are then quite different from

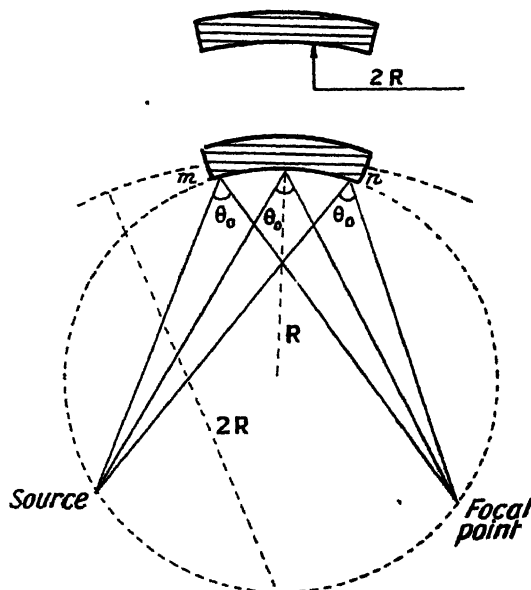


Figure 3. Reflexion by a cut and bent crystal.

those of ordinary diffraction patterns, where either the crystal is immovable but the beam polychromatic, or the crystal is rotated and the beam monochromatic. To obtain a beam which is both monochromatic and intense, we use an apparatus (figure 3) which we shall now describe very briefly (Guinier, 1939 a). The

monochromator is a bent quartz crystal which gives a reflected beam, monochromatic and convergent, with an aperture angle of 2 or 3°. We use a lamella, cut and bent, as described by Johansson (1933), so that the focusing of the reflected beam is theoretically accurate. The x-ray source is the very sharp focus of a tube, with a target parallel to the monochromator crystal.

The diffraction camera is designed in order to use the whole energy of the convergent beam with the greatest efficiency. For instance, to make a Debye-Scherrer pattern, the photographic film and the sample are on the same cylinder as in Seeman-Bohlin's device (figure 4a). Thus the lines are very sharp, what-

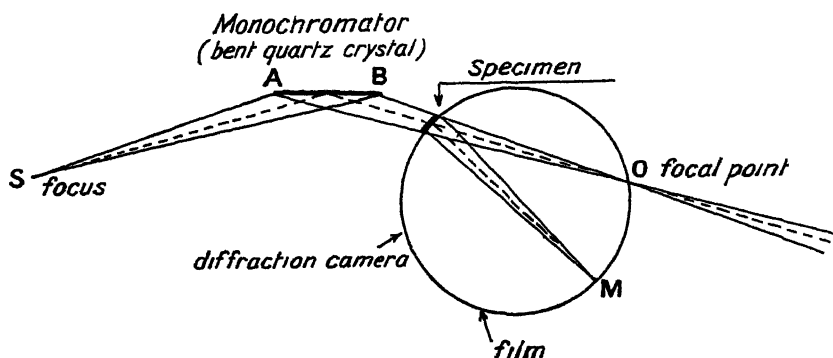


Figure 4a. Device for powder patterns with monochromator

ever the aperture of the incident beam through the sample. Figure 5 gives some examples of such patterns. The time of exposure is the same as with an ordinary Debye-Scherrer camera.

For single crystal patterns it is not sufficient to record the scattered rays only in the equatorial plane. The sample is on the axis of a cylindrical camera placed in such a way that the monochromatic beam converges on a point of the cylindrical film (figure 4b). Focusing is not obtained with mathematical accuracy, but in

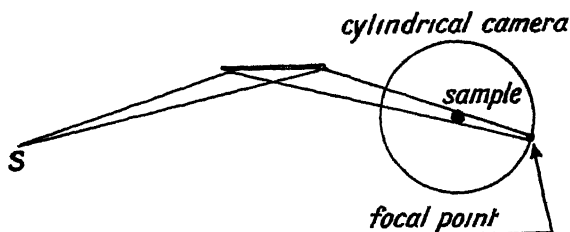


Figure 4b. Device for rotating-crystal patterns.

the central part of the pattern the spots are at least as well defined as in a usual pattern. The whole apparatus can be placed in an evacuated box, in order to avoid scattering by the air.

§3 SOME APPLICATIONS OF THE STUDY OF X-RAY SCATTERING

We may now review some examples of lattice imperfections which produce anomalous scattering. They are of various kinds, but not every lattice imperfection produces observable scattering. One can only study by this method perturbations which extend over a large part of the crystal. For instance, it is

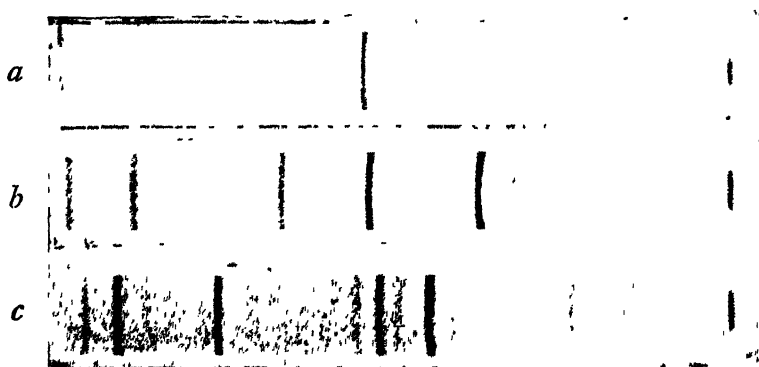


Figure 5 Powder patterns with monochromatic $\text{CuK}\alpha$ radiation.

(a) Nickel foil (exposure, 5 min)

(b) NaCl

(c) Aluminium alloy with a small amount of Al_2Cu

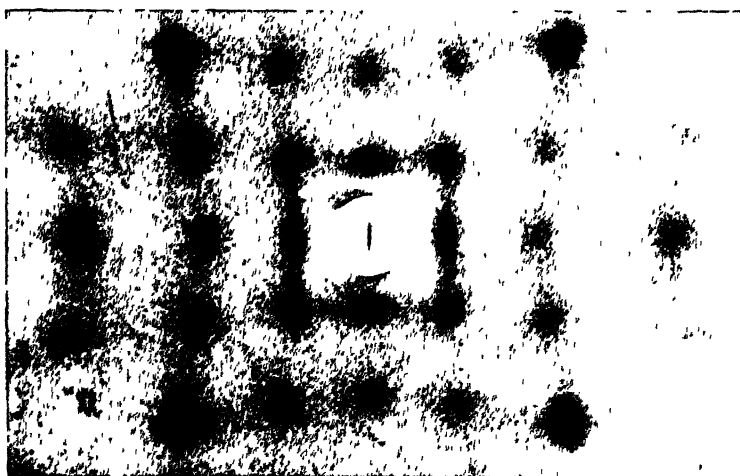


Figure 6. Thermal scattering pattern lamella of KCl parallel to (100) plane, normal to the incident x ray (exposure, 6 hours)

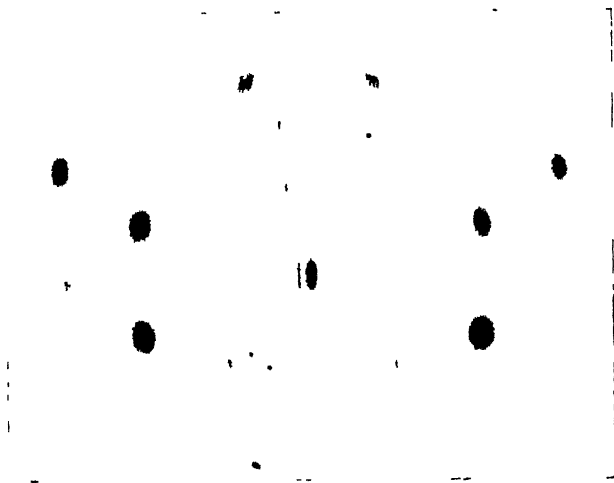


Figure 7 Scattering pattern by an aluminium crystal after a 30% extension.

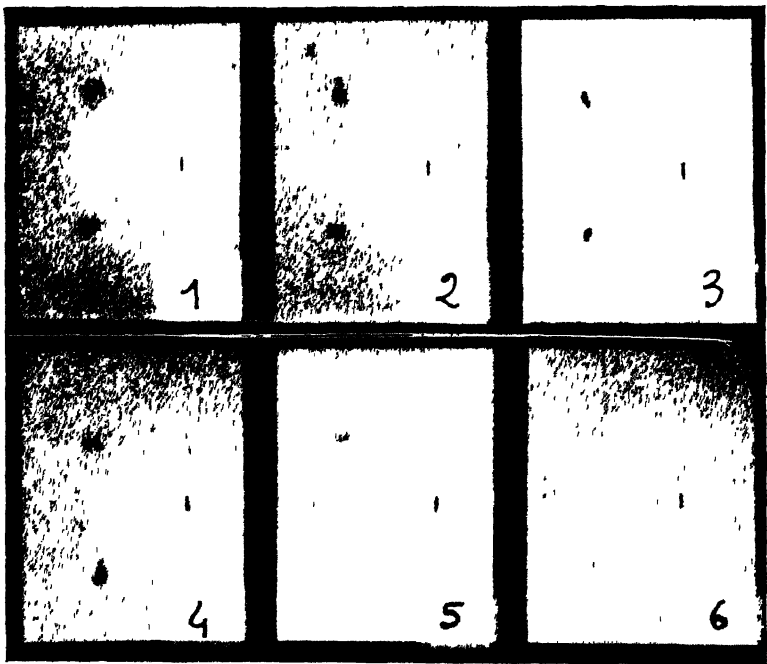


Figure 10 Anomalous scattering by diamond (with $\text{MoK}\alpha$ radiation) The Bragg spots 111 and $1\bar{1}$ appear in photograph no 3 From this position the crystal is rotated around the 001 axis.

- (111) Nodes inside the reflexion sphere { no 1 rotation, $4^\circ 5$).
no 2 rotation, $1^\circ 5$).
- (111) Nodes outside the reflexion sphere { no. 4 rotation, $1^\circ 0$).
no 5 rotation, $3^\circ 5$)
no 6 rotation, $5^\circ 5$)

not possible to distinguish a perfect crystal from a mosaic crystal, because the latter is formed by a collection of small blocks joined together with slight disorientations, each block having a perfect lattice, so that the disturbed regions represent too small a volume of the whole crystal.

Thermal motions of atoms

A first type of lattice imperfection, very important on account of its general occurrence, is the thermal agitation of atoms. Every crystal lattice, at every temperature, presents this kind of imperfection. Thus, on all scattering patterns, the "thermal" scattering is visible to a greater or less extent. This scattering, though observed many years ago by a great number of authors, has only recently been carefully studied. Now there is a considerable amount of literature on this subject.* It is the most important, and the most generally known, of lattice imperfections investigated by x-ray scattering. One of the aims of this paper is to show that there are other various kinds of imperfections revealed by the same method. It is not possible, here, to enlarge upon this very extensive and somewhat difficult subject; we will only reproduce some photographs (figure 6) obtained with the apparatus above described. The interesting feature is that they are obtained with quite short exposures. For the photograph of NaCl, the time of exposure is 6 hours, whilst a similar pattern obtained by Gregg and Gringrich (1941) with a plane crystal monochromator required 100 hours.† The results of the study of thermal scattering are, mainly, the determination of the elastic constants of crystals and the comparison of the propagation velocities of the elastic waves in different directions, these two factors determining the atomic movements in the crystal. Thus, x-ray scattering now becomes an important method for the study of the elastic properties of solids.

Lattice deformations under mechanical stresses

A monocrystalline aluminium wire is progressively extended by traction. Gliding occurs upon one or many sets of (111) planes. It is well known that during this process the crystal undergoes rotations and, in addition, asterisms appear on Laue patterns, because the differences of the orientations of the mosaic blocks grow larger. We attempted to discover some perturbations inside the individual blocks by studying the scattering patterns. The pattern of figure 7 corresponds to a 30% extension of the wire: *it is identical with the pattern obtained before distortion*: the only new feature is the appearance of faint Debye-Scherrer spots in addition to the diffuse spots due to the thermal agitation. This result, though negative, is interesting. One might have imagined that, whilst gliding, the successive glide-planes would be displaced parallel to themselves with some irregularities. If this hypothesis had been true, we should have seen the diffraction pattern of the two-dimensional lattices of aluminium planes (111). This pattern is formed by anomalous spots corresponding to the intersection of the reflexion sphere with the (111) axes of the reciprocal lattice. We have never succeeded in observing any such spots. We think therefore that, whilst gliding, blocks comprising a considerable number of (111) planes move together and

* See, for example, *Proc. Roy. Soc. A*, August 1941.

† Private communication.

remain without internal defects. Gliding takes place upon a few planes far apart: thus the perturbed domains have too small a relative volume to scatter more than a negligible amount of energy. In regions where the atomic disorder is too great, the metal recrystallizes; the newly formed microcrystals give rise to the Bragg spots, visible on the photograph. Thus, during the deformation of the wire, the monocrystal is broken into several crystals, but no lattice deformations of these new crystals can be revealed by the x-ray scattering.

Defects of periodicity in crystals

One-dimensional defects. An example is given by the anomalous scattering of diamonds (Lonsdale, 1941; Raman, 1941; Guinier, 1942 a).

Let us consider a lattice as a set of identical planes parallel and regularly spaced. Each of these planes gives, for all incident rays, a diffraction pattern. On account of the perfectly regular spacing of successive planes, these patterns vanish completely by interference, except for the special directions which correspond to the selective reflections. But, if the interplanar spacings were distributed at random, the diffraction pattern of the whole would be that of a unique plane. What is it? Let us construct the reciprocal lattice of the two-dimensional lattice: the scattering zones in the reciprocal space are the rows normal to the plane, passing through the nodes of the plane reciprocal lattice.

In fact, however, in actual disturbed lattices, the subplanar spaces do not take all possible values but vary over a range which remains near the value of

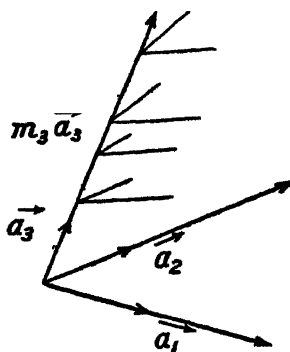


Figure 8 Succession of (001) planes slightly displaced

the perfect lattice; furthermore, sets of normally spaced planes can persist. Consequently, the reciprocal space contains, firstly, the nodes of the reciprocal space of the ideal lattice, corresponding to a strong diffracted intensity, and secondly, zones of low intensity, corresponding to the patterns of the plane lattice, along the rows defined above. The scattered intensity varies along the row in accordance with the distribution of the interplanar spacings. As an example of the interpretation of a scattering pattern, we shall show how a statistical scheme of the irregularities of the perturbed planes may be deduced from the variations of the scattered intensity.

Consider a lattice built upon the three axes $\vec{a}_1, \vec{a}_2, \vec{a}_3$. We suppose the successive (001) planes to be displaced along the direction \vec{a}_3 (figure 8). The

coordinates of any atom (structure factor f) are $\vec{m}_1 \vec{a}_1, \vec{m}_2 \vec{a}_2, \vec{m}_3 \vec{a}_3$, but only m_1 and m_2 must be integers. Let M_1, M_2, M_3 be the number of unit cells along the three axes, in the diffracting body. The scattered intensity is:

$$= f^2 \left[\sum_0^{M_1} \sum_0^{M_2} \sum_0^{M_3} e^{-2\pi i (\vec{b} \cdot \vec{m}_1 \vec{a}_1 + \vec{b} \cdot \vec{m}_2 \vec{a}_2 + \vec{b} \cdot \vec{m}_3 \vec{a}_3)} \right]^2. \quad \dots\dots (3)$$

We put x_1, x_2, x_3 for the coordinates of $\vec{b} = \frac{\vec{S} - \vec{S}_0}{\lambda}$ in the reciprocal space, referred

to the axes of the reciprocal lattice $\vec{A}_1, \vec{A}_2, \vec{A}_3$. Thus,

$$\vec{b} = x_1 \vec{A}_1 + x_2 \vec{A}_2 + x_3 \vec{A}_3.$$

Substituting in equation (3),

$$I = f^2 \sum_0^{M_1} e^{-2\pi i x_1 m_1} \sum_0^{M_2} e^{-2\pi i x_2 m_2} \sum_0^{M_3} e^{-2\pi i x_3 m_3}$$

In the first two factors, m_1 and m_2 are integers: thus they are only different from zero if x_1 and x_2 are integers. This means, as we have stated above, that the vector \vec{b} must terminate on one of the reciprocal lattice rows parallel to \vec{A}_3 . If this condition is fulfilled, the intensity is

$$I = f^2 M_1^2 M_2^2 \left| \sum_0^{M_3} e^{-2\pi i x_3 m} \right|^2$$

or

$$\begin{aligned} I &= f^2 M_1^2 M_2^2 \sum_{m_3} \sum_{m_3'} e^{-2\pi i x_3 (m_3 - m_3')} \\ &= 2f^2 M_1^2 M_2^2 \sum \sum \cos 2\pi x_3 \Delta m_3 m_3'. \end{aligned} \quad \dots\dots (4)$$

$\Delta m_3 m_3'$ represents the absolute value of the difference in ordinate between the two reticular planes m_3 and m_3' , the unit of length being $|\vec{a}_3|$. On an axis OZ, we plot from the origin O a segment equal to $\Delta m_3 m_3'$ for each pair of reticular planes of the actual crystal. The total number of points is M_3^2 . Let $f(x)$ be the linear density at abscissa x , i.e. $f(x)dx$ is the number of points lying between x and $x+dx$.

The sum (4) can be replaced by the integral

$$I = 2f^2 M_1^2 M_2^2 \int_0^\infty f(x) \cos 2\pi x_3 x dx. \quad \dots\dots (5)$$

This formula determines the scattered intensity I along a row of the reciprocal lattice, when the spacing of the planes, or $f(x)$, is known. Conversely, after inversion, this Fourier integral gives $f(x)$ as a function of I , that is to say, as a function of the scattered intensity which is measured in the experiment. In this method *the amplitude of the scattered radiation does not appear*: the difficulty encountered in the general method (formula 2) is avoided. $f(x)$ is proportional to

$$\int_0^\infty I(x_3) \cos 2\pi x_3 x dx_3. \quad \dots\dots (6)$$

Thus photometric measurements of scattering patterns lead to an accurate knowledge of the perturbations of reticular planes by means of the function $f(x)$.

As an example, let us assume that the spacing of two neighbouring planes remains close to the theoretical value but that the perturbations are accumulated from cell to cell, so that the difference between the ideal and actual spacings of

two planes grows larger with increasing distance between them. What is the form of $f(z)$? At first, M_3 points meet at the origin. If the crystal were perfect, the $M_3 - 1$ spacings between consecutive planes would give $M_3 - 1$ points at $z = 1$: actually, these $M_3 - 1$ points are grouped about this value. The curve $f(z)$ shows a narrow maximum about $z = 1$. The spacings between every second plane are grouped about $z = 2$, but they diverge rather more: the second maximum is wider and lower than the first, etc. When z is big enough, $f(z)$ tends towards a constant value. There is no long-range order. Figure 9 gives a schematic function of this type. The corresponding variation of I , according to the formula (5), is plotted below. The scattered intensity presents maxima for $z = 1, 2$, etc., that is to say, about the nodes of the reciprocal lattice of the perfect crystal. The scattering is chiefly observed *around the node of index 1* and is *weaker for greater indices*. Furthermore, it is symmetrical on each side of the node.

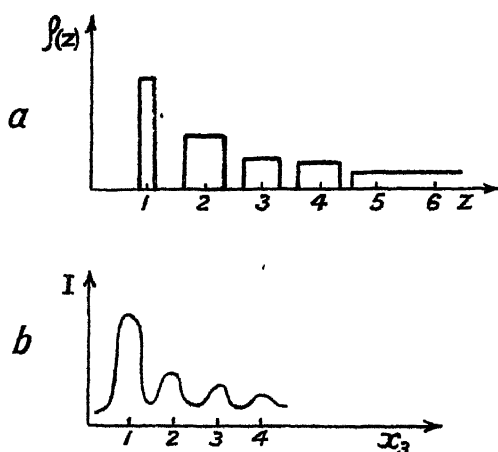


Figure 9. (a) Successive interplanar spacings, (b) diffracted intensity plotted against the ordinate along the rows (001).

Scattering of this type has been effectively observed in the case of diamond by several physicists. When the crystal is so directed that a (111) node is near to the reflexion sphere, three faint but distinct spots are visible on the pattern (figure 10) around the point where the (111) diffraction spot would appear. They become fainter and diverge from each other as the distance from the node to the sphere increases. These three spots correspond to the intersection of the sphere with the three rows parallel to the cubic axes passing through the (111) node. From the above considerations we conclude that diamonds contain regions where the planes parallel to the faces of the cubic cell are irregularly displaced; and since the scattering is localized near the reciprocal lattice nodes, and mainly near the 111 nodes, the deformations must be of the general type described above: this conclusion is the direct and geometrical result of the experiment.

What can be the origin of these structure irregularities? They seem to be peculiar to the diamond, and especially to one of the two kinds of diamond (type II). Friedel (1924) has shown that there is an allotropic transformation of the diamond at a very high temperature. This second form of unknown

structure is not stable at ordinary temperatures. Let us assume that this second form is produced from diamond crystals by moving (100) planes without perturbations of the atoms in these planes. Then the disturbed regions of diamonds revealed by anomalous scattering would appear as regions where the allotropic transformation to the diamond form is stopped at an intermediate stage. In the study of some alloys we encounter such phases with imperfectly periodic structure, intermediate between two well-known crystalline forms. In the case of the diamond, one of the lattices is unknown, so that the explanation is still only hypothetical.

Two-dimensional defects: example of fibrous crystals

Let us consider the crystalline lattice as a set of parallel rows and let us suppose that, without the disposition of atoms along these rows being disturbed, they are slightly displaced by irregular translations.

In this case it is the diffraction pattern of the *one-dimensional lattice* of the fibres which can be observed. These fibres being normal to the monochromatic incident x rays, this pattern is formed by the well-known "layer lines" of the fibrous crystal patterns. In the perfect crystal patterns, these layer lines are only made visible by the succession of discrete diffraction spots; but, for imperfect crystals, the latter lines are formed by continuous streaks. This phenomenon is observed with some fibrous substances, as in the case of *asbestos*. Figure 11 is a pattern of chrysotile, obtained with $\text{CuK}\alpha$ radiation. It must be noticed how essential is the use of pure monochromatic radiation, for the continuous spectrum would also produce streaks on the central line. A striking feature is that the intensity of the streak is not constant along the whole lines. Streaks start from some diffraction spots, but not from all spots, and on one side only: they are directed towards large diffraction angles. This proves that the perturbations of the fibre are not distributed at random. Aruja, who has also obtained these patterns with chrysotile, thinks that complete sheets of fibres are displaced in some directions. His hypothesis is verified by the photometric measurements of the streaks, and this work is a good example of the accurate information of real crystal structures obtainable by careful study of diffuse scattering.

§ 4. ARRANGEMENT OF THE ATOMS IN A SOLID-SOLUTION CRYSTAL

Order-disorder transformation

A crystal of an ideal solid solution is to be imagined as formed by a perfectly regular lattice, at the nodes of which two kinds of atoms are distributed at random. The periodicity is then rigorous for the *positions* of atoms, but not for their diffracting power (or structure factor): this latter lack of periodicity is a cause of scattering. Let a crystal contain N atoms: $p_A N$ of these are of type A and have a structure factor f_A , and $p_B N$ are atoms of type B and have a structure factor f_B .

Von Laue has shown that the solid-solution crystal gives rise, on one hand, to the diffraction pattern of a crystal with the same lattice, all the nodes of which are occupied by a unique fictitious atom of structure factor $p_A f_A + p_B f_B$, and, on the other hand, to a diffuse scattering, the intensity of which is

$$I = N p_A p_B (f_A - f_B)^2.$$

This formula depends only on the scattering angle, but not on the crystal orientation; it shows a maximum in the direction of the incident beam, as in the case of a gas. Laue's calculations have not previously, I believe, been experimentally verified. We have undertaken this verification and have chosen, for this purpose, the equiatomic gold-silver alloy, because of the advantages that the two atoms have very different structure factors but almost equal radii: thus, for any atomic distribution, the lattice deformations remain very small.

The patterns we have obtained are in opposition to Laue's theory. The scattering presents no maximum at the centre of the pattern. The lower curve of figure 12 represents the scattering of a sample formed by two distinct sheets

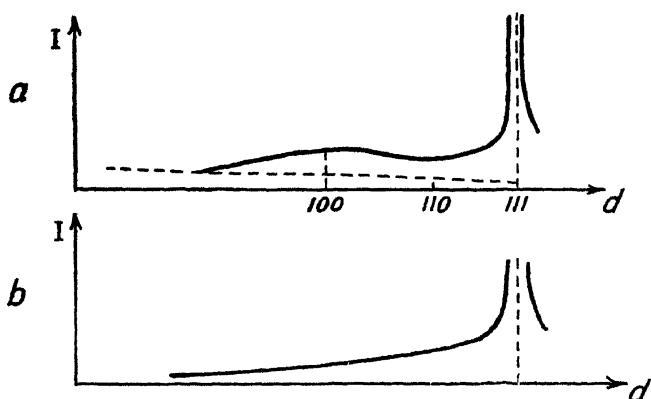


Figure 12 Intensity of scattered radiation from an Au-Ag alloy (a) Au-Ag alloy, (b) metals Au + Ag

of pure gold and pure silver. It is mainly due to the thermal effect. The upper curve represents the scattering by a sheet of the Au-Ag alloy, and the dotted curve the scattered intensity according to the Laue's formula. If the latter were correct, the observed curve would be nearly the sum of the lower and the dotted one. But it is not true. There appears a faint and wide ring, the maximum of which lies at the place of the 100 line of the crystal. The intensity of this ring is of the same order of magnitude as the maximum intensity given by Laue's formula.

Hence Laue's hypothesis is not valid, i.e. the *two kinds of atoms are not distributed at random*. Probably a silver atom, for instance, tends to be surrounded by gold atoms and vice versa. In a face-centred cubic lattice of parameter a , an atom, say an Ag atom, has 12 neighbours at a distance of $a/\sqrt{2}$, then 8 at a , and so on. If the Ag and Au atoms were distributed at random there would be (6 Au and 6 Ag) at $a/\sqrt{2}$, (4 Au and 4 Ag) at a , etc. Actually there must be fewer Ag atoms among the first neighbours and more in the second, that is to say, there is, between two neighbouring Ag atoms, a preferred spacing, as between two molecules in a liquid. The effects upon the scattering pattern will be similar; a ring appears at an angle which corresponds to this preferred distance in the case considered it is the distance a , that is to say, the (100) spacing, which explains our experimental results.

The scattering pattern seems to show that perfect disorder in a solid solution does not exist: however, no ordered super-lattice has ever been found in Au-Ag

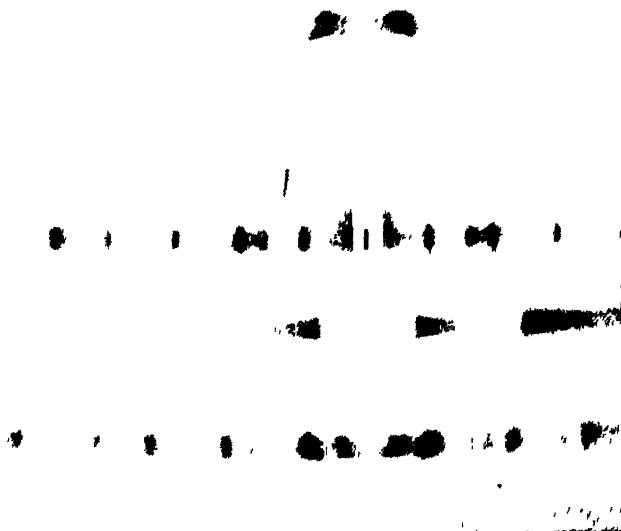


Figure 11. Scattering pattern of chrysotyle ($\text{CuK}\alpha$ radiation)

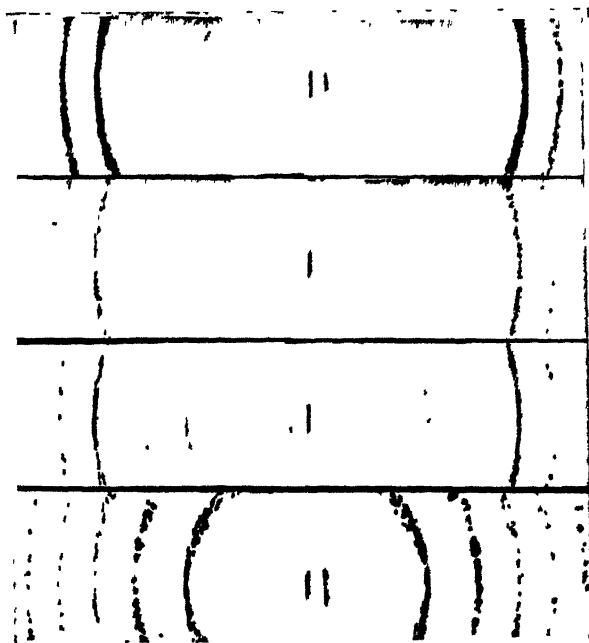


Figure 14 Scattering patterns of Au-Cu_3 alloys.

Scattering patterns of annealed samples

- (1) Annealed, 2 hours at 100°C .
- (2) Annealed, 2 hours at 250°C .
- (3) Annealed, 2 hours at 300°C .
- (4) Annealed, 2 hours at 350°C .

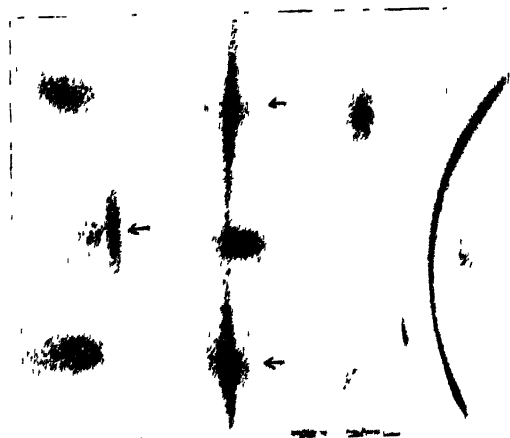


Figure 17 Scattering pattern of AuCu_3 crystal (surrounded by microcrystals) The arrows show the scattering zones around normal nodes

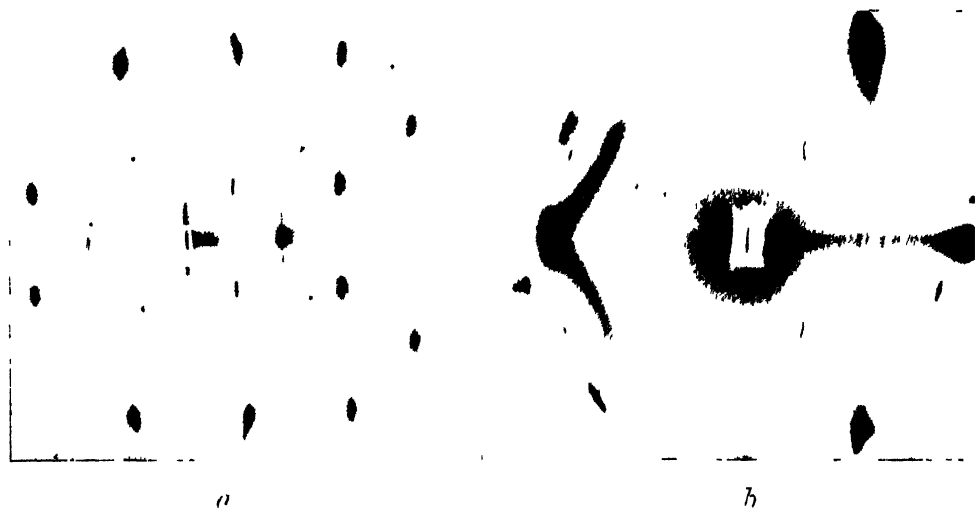


Figure 18 Scattering patterns of age-hardened alloys (a) Al-Cu alloy, (b) Cu-Be alloy

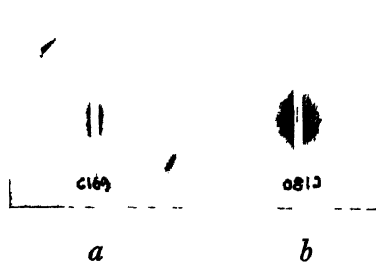


Figure 19. Small-angle scattering by charcoal (a) before activation; (b) after activation.

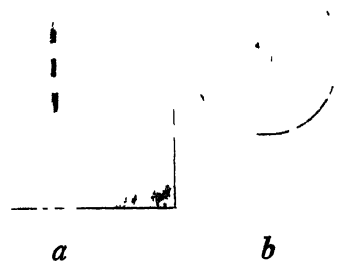


Figure 20. Small-angle scattering by age-hardened alloys (a) Al-Cu; (b) Al-Zn.

alloys. But, even in this case, the nature of an atom has some influence on that of its neighbours: there exists what Bethe has called *short-range order*, but to a small extent. We have also investigated an alloy which can take an ordered form, namely AuCu_3 . In its disordered state (quenched) one finds also on its scattering pattern a broad ring, as in the case of Au-Ag , but notably more intense (figure 13). Even in this case there is no perfectly disordered state at room

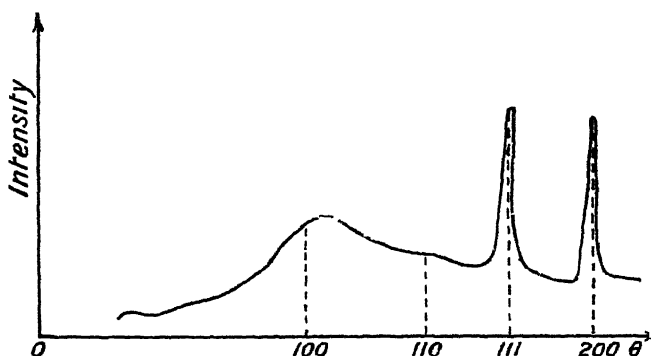


Figure 13.—Scattering patterns of Au-Cu_3 alloys. Scattered intensity from a quenched sample.

temperature, and the degree of short-range order is greater than in the case of Ag-Au . Experiments are now being carried out to see whether the same applies above the critical temperature.

Let us consider next the transformation from the disordered to the ordered state: during the annealing of the alloy this change is quite progressive on the pattern (figure 14), the broad ring splits into two distinct lines, the one at the

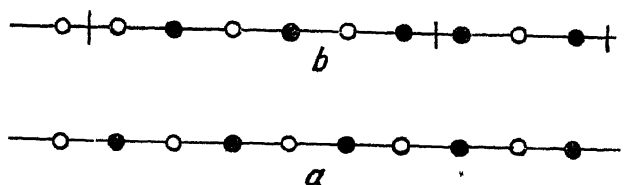


Figure 15. (a) Perfect order, (b) partial order with out-of-phase domains.

(100) position and the other at the (110). These two rings, at first diffuse, become narrower and more intense, at the same time, rings of higher indices make their appearance. Sykes and Jones (1936), studying these phases of transformation, concluded that the lattice was formed by juxtaposition of small domains which are ordered, but out of phase. Let us give an example in the schematic case of a one-dimensional lattice and an equiatomic mixture of A and B. The lower line of figure 15 shows the ordered state and the upper line a partially ordered state, with several elementary ordered domains; thus, the normal lines are not modified and the superstructure lines are broadened by the well-known effect of small-size crystal. With further ordering, the elementary domains grow larger, and the lines become more distinct. We think that such a representation is far too simplified when applied to the earlier stages of the

ordering process. A more accurate theory of the patterns will link the scattering with the parameter measuring the short-range order. Let us consider again the case of a one-dimensional lattice, a row of N equally spaced atoms with a period 1, which comprises $N/2$ atoms A and $N/2$ atoms B. The short-range order can be defined by the probability, τ , that any atom may have a neighbour of the same nature; τ varies from $\frac{1}{2}$ to 0 when the lattice passes from the perfectly disordered to the perfectly ordered state. The scattering pattern can be calculated throughout as a function of τ (figure 16). For $\tau = \frac{1}{2}$ one finds the lines of a disordered

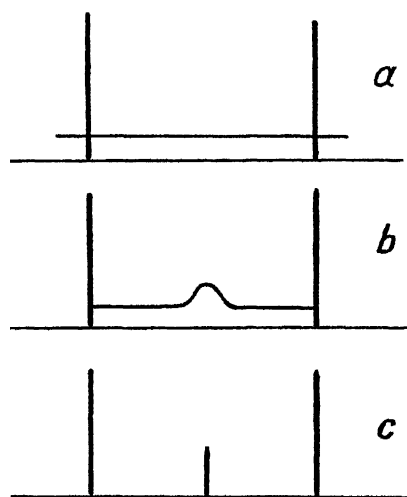


Figure 16. Scheme of scattered and diffracted intensity
(a) Perfect disorder, (b) partial order, (c) perfect order

lattice (period 1) and a continuous background given by Laue's formula. As τ tends to 0, the scattering intensity presents maxima, midway between successive diffraction lines. These maxima become higher and narrower as τ decreases. When τ equals 0, the superstructure lines take the place of what were scattering maxima. These considerations conform with the observed phenomena in the case of a three-dimensional lattice. Thus, the scattering patterns are proved to be a new experimental tool for the quantitative study of order in solid solutions.

Up to the present, we have presumed that the solid-solution lattice, ordered or not, was perfectly regular. But this is probably not generally true. The irregular distribution of atoms of different volumes produces lattice deformation. This is shown by the scattering pattern of a monocrystal of AuCu_3 alloy. As has already been mentioned, the scattering zones around the superstructure nodes are due to a partial order: but no distribution of atoms can explain scattering zones about the normal lattice nodes (see figure 17). These are produced by lattice distortion. It is well known that ordering is often accompanied by the transformation of the cubic lattice into a tetragonal one. This transformation is also progressive, and in the intermediate state the lattice is no longer perfectly periodic, leading to the appearance of anomalous scattering.

Thus, in the study of order-disorder transformation, x-ray scattering has revealed two new features: the short-range order in the so-called "disordered state", and the geometrical lattice deformations.

Age-hardening

Another very interesting kind of solid solution is one containing an excess of one kind of atom, and which is capable of age-hardening. The dissolved atoms, in excess, tend to form a precipitate with a distinct crystalline structure. But before the appearance of the precipitate, the properties of the alloy, mechanical, electrical, etc., have been changed, although microscopic examination does not reveal any crystal modification. This is the phenomenon of age-hardening. The scattering patterns give a very detailed view of the process of precipitation. The first results, in that field, were obtained in 1938 for Al-Cu alloys independently by Preston and by Guinier. Since then, other alloys have been studied by different authors: I have worked especially (1942, 3, 4) on Al-Ag, Al-Zn, Cu-Be alloys. The results of these studies, too, are of great theoretical and practical value in this section of metallography. It is, of course, very interesting to know the actual nature of those lattice perturbations which increase the mechanical strength of metals. Figure 18 shows some examples of scattering patterns corresponding to successive stages of the age-hardening process.

§ 5 SCATTERING BY SMALL PARTICLES

We shall consider finally one more kind of lattice imperfection. The directions of the selective reflection are only defined if the crystal is large enough. When it is very small (dimension less than 500 Å.), each node of the reciprocal lattice is replaced by a small region surrounding it: accordingly, the lines of a Debye-Scherrer pattern are broadened. This effect has been extensively studied, and it is possible to determine the size of the particle from the width of the Debye-Scherrer lines. These studies can be carried out with an ordinary camera and polychromatic radiation.

The origin of the reciprocal lattice is also a node—it is, then, also surrounded by a small scattering zone, when the crystal is very small. That is to say, a sample formed by small crystals produces some *scattering at very low angles*. The study of this central scattering is more difficult than that of the broadening of the Debye-Scherrer lines: the use of a monochromatic radiation is advantageous, and the bent crystal monochromator is particularly convenient (Guinier, 1939 b, 1943 b). But, on the other hand, the central scattering gives more useful information: it is independent of the internal structure of the particle—it depends only on the size and the shape of the particle. It follows that the method can be used also for *amorphous substances*. The central scattering is characteristic of the divided state of matter. For instance, figure 19 shows two photographs relating to similar samples of charcoal, but the second is activated charcoal. The central spot is broad (3 or 4°), which proves that, after the activation treatment, the matter is in a very finely divided state.

In the age-hardening alloys, the atoms of the solute, before precipitation, gather into small clusters dispersed in the crystal of solid solution. These clusters produce central scattering, and then the size and shape of these sub-microscopical clusters can be determined. Figure 20 gives two examples, the first concerning Al-Cu (Guinier, 1942 a) and the second Al-Zn (Guinier, 1943 a).

§ 6. CONCLUSION

The Bragg-Laue theory and the x-ray diffraction patterns relate only to the ideal geometric lattice. In our knowledge of the structure of the solid state

they form a first approximation. Next, physicists try to build a model of the actual crystal and are more and more concerned with lattice imperfections. The disturbed lattice is very often encountered, especially in the case of alloys. The study of the diffuse scattering, produced by some kinds of lattice imperfections, is in many cases a very useful experimental method. In this brief review we have attempted above all to show how varied are the fields of application of the scattering patterns.

REFERENCES

- EWALD, P. P., 1940. *Proc. Phys. Soc.* 52, 167.
 FRIEDEL, 1914 *Bull. Soc. franc. Mine* 47, 60.
 GREGG, R. Q., and GRINGRICH, N. S., 1941. *Phys. Rev.* 59, 619.
 GUINIER, A., 1938 *C. R. Acad. Sci., Paris*, 206, 1972, 1939a. *Ann. Chim. (Phys.)* 12, 161; 1939b. *Ibid.* 12, 161; 1942a. *C. R. Acad. Sci., Paris*, 215, 114; 1942b. *J. Phys. Radium*, 8, 124; 1943a. *Metaux et Corrosion*, 18, 209; 1943b. *J. Chim. Phys.*, 40, 133; 1944 *Rev. Metall.* p. 1.
 JOHANSSON, T., 1933 *Z. Phys.* 82, 507.
 LONSDALE, K. and SMITH, 1941. *Phys. Rev.* 61, 617.
 PRESTON, G. D., 1938 *Proc. Roy. Soc. A*, 167, 526.
 RAMAN, C. V. and NILAKANTAN, P., 1941. *Phys. Rev.* 60, 63.
 SYKES, C. and JONES, F. W., 1936. *Proc. Roy. Soc. A*, 157, 213.

PERMEATION AND SORPTION OF WATER VAPOUR IN VARNISH FILMS*

BY A. MORRIS THOMAS AND W. L. GENT,
Perivale, Middlesex

MS received 22 February 1945

ABSTRACT An experimental investigation on the moisture permeability and sorption of detached varnish and polystyrene films is described. Details for preparation of samples and methods of test, under various conditions of vapour pressure and temperature, are given. Both the permeability and sorption of the films have been determined at constant moisture concentrations, that is, by using streams of moist air, the amounts permeating and the amounts sorbed being directly weighed. In addition to moisture permeability, some measurements of that of hydrogen and carbon dioxide are reported.

Consideration of the experimental results shows that they can be interpreted by assuming that the diffusion process is analogous to that of substances in solution. There is insufficient evidence to show the effect of vapour pressure and temperature on the moisture permeability, but the results for carbon dioxide and hydrogen indicate that the diffusion process is dependent on the microphysical properties of the film. The effect of structure on diffusion is discussed, and the relations of the results to the problems of electrical insulation are considered.

§ 1. INTRODUCTION

EXPOSURE of insulating materials to water or its vapour is almost always detrimental to their electrical properties. Cellulosic materials, extensively used owing to their mechanical properties, are badly affected, and so are customarily protected by impregnation with hydrophobic filling waxes, compounds or varnishes. These various organic materials, however, exhibit widely different sorptive and permeability characteristics which it is important

* Based on Report A/T92 of the British Electrical and Allied Industries Research Association.

to investigate with a view both to practical applications and to elucidating the fundamental factors involved.

These considerations apply equally to any form of electrical insulating material, and the practical aims of research in this field are then, first, to provide data for estimation of the amounts of moisture likely to enter insulated electrical equipment under any condition of humidity and temperature, and secondly, by relation of these properties to the structure of the insulating materials, which determines their electrical and mechanical properties, to provide information which will lead to improvement of the materials.

The experimental problem is threefold: it is necessary to find the rate at which moisture penetrates into the insulating material, that is, the rate of sorption; to determine the amount of moisture sorbed when the rate of sorption is zero, that is, when equilibrium is reached; and to measure the rate at which moisture passes through the material when sorption equilibrium is attained. Two series of experiments are, therefore, covered by this report, namely, those concerned with sorption of moisture in the material before equilibrium is reached, and those in which the movement of moisture at equilibrium is studied. These are the sorption and permeability tests respectively. In all cases constant humidity and temperature are assumed.

Before the beginning of the research no experimental methods of sufficient accuracy were available; these have been developed. Certain tests for moisture permeability had been devised and widely used, especially in America, but although giving self-consistent results they were of little use where the aim was to provide data applicable under all conditions. They are, however, simple methods of comparing permeabilities under the same conditions.

The permeability to hydrogen and carbon dioxide has been measured for certain materials to ascertain how the penetration of gases, such as these, differ from that of a polar vapour such as water. Although not of any immediate practical importance, the results give some insight into the mechanism of permeation and allow definite conclusions as to its nature to be made. It might be supposed that the mechanism is one of simple infiltration of the permeating substance through minute channels in the material, but the evidence suggests that it is a molecular process and closely connected with sorption. The same molecular properties which determine the electrical and mechanical behaviour of these substances also determine the permeability and sorptive capacities. The theoretical considerations are concerned with the elaboration of this view.

So far, the materials used have been confined to representative types of baking varnishes and polystyrene, both as free films; but although the numerical results apply only to these substances, the general arguments advanced are valid for any organic insulating material of the type considered.

§ 2. DEFINITIONS AND FORMAL THEORY

Fick (1855) suggested that the diffusion of substances in solution could be represented by the two equations

$$\bar{P} = -D \frac{\partial c}{\partial x} \quad \dots\dots(1)$$

$$\text{and} \quad \frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} \quad \dots\dots(2)$$

The first applies to the equilibrium, or stationary, state, where the concentration gradient $\partial c/\partial x$ at any given point is constant. Equation (2) describes the non-stationary state before equilibrium is reached, the concentration gradient being a function of the x -coordinate, which is in the direction of movement of the diffusing substance. \bar{P} , the constant rate of transfer across the concentration difference, is termed the permeability, and D is the diffusion constant.

Wroblewski (1879) suggested that these equations could be applied to the movement of water as vapour in solid materials, and this is borne out by experimental results. It will be shown that a hydrodynamic interpretation of the motion of the permeating substance, that is, permeation through fine channels or micro-pores, is inadmissible for substances such as were used in the present investigation.

On the assumption that D is not a function of c or t , equation (1) may be integrated

$$\bar{P} = D \frac{(C_1 - C_2)}{d} \quad \dots\dots(3)$$

where the C_1 and C_2 are the boundary concentrations inside a slab whose thickness is d . The permeability is a constant for the particular conditions, and, to make it refer to the material, a *permeability constant* is defined as the rate of transfer of mass across unit area and through unit thickness under the influence of unit pressure, or concentration, difference

$$P = \frac{m}{At} \frac{d}{(p_1 - p_2)}, \quad \dots\dots(4)$$

$$P' = \frac{m}{At} \frac{d}{(c_1 - c_2)}. \quad \dots\dots(5)$$

Whereas for the solution system there is a continuous phase, so that $C_1 = c_1$, $C_2 = c_2$, the system vapour-solid is discontinuous at the surfaces and $C \neq c$. The relation that these quantities bear to one another depends on the sorptive properties of the material under consideration,* and if it is assumed that Henry's law may be applied, that there is a linear relation between the external concentration or vapour pressure and the internal equilibrium concentration, then

$$C = Sp \quad \dots\dots(6)$$

or

$$C = S'c, \quad \dots\dots(7)$$

where S and S' are the sorption constants for the material. On writing equation (3) in the form of equation (4),

$$\bar{D} = \frac{m}{At} \frac{d}{(C_1 - C_2)}, \quad \dots\dots(8)$$

it is seen that the relation between the sorption, diffusion and permeability constants for the material is

$$P = SD. \quad \dots\dots(9)$$

* Because of the attraction of the solid for water molecules, the concentration inside the material is greater than that in the vapour phase. It follows that the concentration difference within the material is also greater. Consequently, the constants derived from Fick's equation will be different by a factor which is the ratio of the concentrations in the solid and vapour phases. The coefficient calculated from the concentrations in the vapour phase is the permeability constant, that derived from the internal concentration is the diffusion constant, and is analogous to the coefficient for diffusion in solution,

By inspection of equation (2) it is clear that its solution will yield a kinetic relation between C and t involving D , and a solution of particular interest is for the sorption at both faces of a detached film in an atmosphere containing water vapour at a particular partial pressure. In such a case the boundary conditions are

$$C_1 \text{ (at } x=0) = C_2 \text{ (at } x=d) \text{ for all values of } t, \\ C_0 = 0 \text{ at } t=0,$$

C_1 being defined by equation (6), where p is the ambient vapour pressure, and C_0 the initial concentration in the film. Equation (2) then gives the solution

$$C_t = C_1 \left\{ 1 - \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{1}{2n+1} \sin \frac{(2n+1)\pi x}{d} \exp - \left[\frac{D(2n+1)^2 \pi^2 t}{d^2} \right] \right\},$$

where n is an integer. The total amount sorbed by the film is

$$m = \int_0^d C_1(x, t) dx,$$

which on substitution becomes

$$m = m_s \left\{ 1 - \frac{8}{\pi^2} \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} \exp \left[- \frac{D(2n+1)^2 \pi^2 t}{d^2} \right] \right\}. \quad \dots\dots (10)$$

The series converges rapidly and, when m/m_s is greater than 0.4, equation (10) may be written as

$$\frac{D\pi^2 t}{d^2} = \ln \frac{m_s}{(m_s - m)} + \ln \frac{8}{\pi^2} \quad \dots\dots (11)$$

where m is the mass of water sorbed in time t and m_s is the value at equilibrium. From equation (11), D may be calculated if the rate of sorption be determined, and since in the same experiment S will be found from the equilibrium value of m , a comparison of the value of P from equation (9) and by direct experiment is possible.

The assumptions implicit in the derivation of the three constants and their relationships are given below.

1. The movement of water vapour in non-porous solids is similar, formally, to that of substances in solution.
2. The amount of water vapour sorbed is directly proportional to the ambient vapour pressure.
3. The diffusion constant is not a function of the amount of water sorbed.
4. The time required for the surface to come into equilibrium with the ambient vapour is vanishingly small.

For real substances these assumptions are not necessarily valid, and experimental results should indicate how far they may be used, and (by consideration of the deviations) the real nature of the processes.

§ 3. MATERIALS

The compositions of the varnishes used in the investigation are given in table 1. The films were prepared without alteration of the viscosity of the varnish, as received from the maker.

Four polystyrene films were investigated, details of which are given in table 2, three of the films being of different manufacture.

§4. EXPERIMENTAL METHODS

Although the investigation was primarily concerned with varnishes, which are normally used as thin films or as impregnants, other materials may be studied advantageously in this manner. For films up to 0.015 cm. in thickness, equilibrium is established within a few hours for both permeation and sorption

Preparation of varnish films

In many researches, e.g. Gettens (1932), Gettens and Bigelow (1933), Payne and Gardner (1937), Payne (1939), the varnish has been supported on paper or on glass cloth (Payne, 1939). Although such a method is useful for comparative

Table 1. Compositions of varnishes *

Varnish No	Components	Per cent	Solvent	Per cent
1	Tung oil Phenolic resin (100%)	40.6 12.0	Mixed petroleum distillate and terpene	47.4
2	Gilsonite Boiled linseed oil	33.6 13.2	White spirit	53.2
3	Spirit soluble phenolic resin (100%)	38.4	Methylated spirit Acetone	53.9 7.7
510	Tung oil Phenolic resin (100%) Drier	41.2 11.6 1.2	Mixed petroleum distillate and terpene	46.0
602	1 part of No 510 and 1 part of : Congo Copal ester . 53% Boiled linseed oil . 17% Solvent . 25.8% Drier . 4.2%			

* Varnish 1 also contained drier of unspecified amount. The suppliers of the varnish state that Nos. 1 and 510 were identical in composition

The baking times were 2 hours per coat for varnishes Nos 1, 2, 510 and 602, and 3 hours per coat for varnish No 3. The baking temperature was 105° c in all cases

Table 2. Details of polystyrene films

Film No.	Thickness (cm)	Description
1358	0.0025	Unplasticized
1358A	0.0075	Unplasticized
1359	0.0125	Unplasticized
1360	0.0025	—

tests, it is of little use for determination of absolute values of the constants, as the thickness of the varnish film and effect of the base material are always indefinite. When constants referring only to the varnish are required, detached films must be used.

The first necessity for precise measurements is that the films be uniform in thickness, and it is, therefore, not possible to prepare them by brushing or spraying. In general, two methods have been used first, spinning on a horizontal plate (Lishmund and Siddle, 1941), or secondly, spreading by means of a doctor blade moving at a fixed height above the base on which the film is to be formed. The second method has been employed in this investigation.

The second necessity is that the film when formed should be easily removed from the base material. Most investigators have used tinned-iron plates from which the film may be stripped, but it has been found more advantageous to use 3-mil tin-foil which is dissolved off by mercury.

The apparatus used in this investigation is shown in figure 1. The doctor blade is attached to two differential screws, with large knurled heads, which are

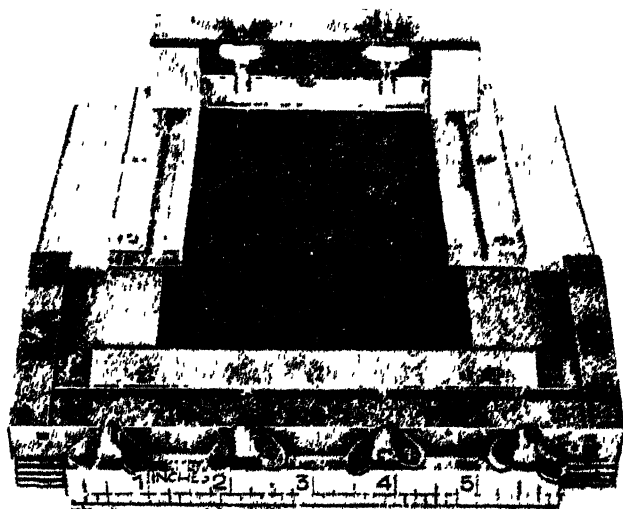


Figure 1. Apparatus used for the preparation of varnish film samples

supported in an iron angle at the ends of which are steel runners sliding in steel grooves clamped to the brass base plate. It is on this plate that the tin-foil is stretched by application of a slight tension to the clamps at each end. The gap beneath the blade edge is made parallel by adjustment of the differential screws and then set to about 0.005 cm. A small quantity of the varnish is poured on the foil and the blade carrier moved slowly along the guides so that the excess of varnish is pushed in front. Just before the edge of the base plate is reached, the carrier is reversed and the movement repeated in the opposite direction. The to-and-fro motion is carried out three times, and the excess of varnish pushed over the sloping edge of the base plate and removed. The foil is released and baked for the appropriate time. Further coats of varnish are added by replacing the foil and repeating the process. To avoid dust, which gives rise to pinholes and irregularities, the whole apparatus is covered by an enclosure with curtain-sides and roof of cellophane. During baking, the film is covered with a shield of the same material. The films are substantially uniform over an area of 10 cm. \times 15 cm. Most of the films used were manipulated with ease, but those of varnishes 2 and 3 presented special difficulty.

Films of varnish No. 2 are fragile, due to the shortness of the mixture, and further, at temperatures above 40°C. the viscosity decreases so much that the film becomes semi-liquid. For this reason no values of permeability constant were obtained at 50°C, and those at 40°C. were subject to some uncertainty. The tendency to fluidity is due to lack of polymerization, the film being produced by desolvation. This mode of formation also gave rise to difficulty in the sealing of rings into the permeability apparatus. Rubber rings were used normally, but with this varnish it was found that after three days the rubber had sorbed all of the oil and left only a powder of gilsonite, the centre of the film becoming separated. The effect was overcome by using cellophane masking rings over the area of contact with the rubber.

Fragility of films of varnish No. 3 is due to their brittle nature. The unevenness of pressure obtained with rubber sealing rings caused the films to shatter like glass. To avoid this, rings of waxed filter paper were used, the whole permeability cell being warmed to 60°C. to ensure complete sealing. It is not certain that the films of varnish No. 3 were homogeneous, for polymerization is complete after baking of each coat, so that interdiffusion with the added liquid varnish does not occur.

In all cases the thickness of a film was measured by a dial micrometer reading to 0.0001 cm. A number of readings were taken over the area under test. The softer films presented some difficulty as the impact of the micrometer foot caused compression, giving a lower value than the true thickness. For this reason a circular plate of 1 cm.² area was preferred to a spherical point.

The apparent thicknesses of polystyrene films when measured with the micrometer led to values of density <1, and it was necessary to determine the density directly. It seems probable that the surface is corrugated, and that the crests alone contribute to the dimension.

Production of standard humidities

It is of the first importance both in the permeability and sorption experiments to maintain a constant, standard humidity. To accomplish this, two methods are available for flowing gas streams; the first is the use of saturated salt-solutions which give a known vapour pressure, and the second is to mix such volumes of saturated and dry gas that the resulting mixture has the requisite humidity. The first and simpler method has been used.

Gravimetric checks were made on the humidity of the gas streams, but, in general, the values from *International Critical Tables* were assumed to be correct. The salts which have been used are listed in table 3.

Table 3. Humidities of saturated solutions

Salt	Relative humidity (%)	Variation with temperature
Potassium sulphate	96	None
Potassium tartrate	70	None
Cobalt chloride	67.4, 20° C. 56.7, 40° C	Linear

Measurement of permeability

A variety of forms of apparatus have been described by several authors and comprehensive reviews of the subject have been written by Lishmund and Siddle (1941) and by Carson (1937). To obtain significant results it is necessary that the vapour pressures at the surfaces of the film under test should be of definite and known values. In the majority of the methods used, where the film is exposed over a solution or dehydrating agent without mixing of the air other than by convection, it is probable that the vapour pressure is by no means equal to its supposed value (see, for example, Abrams and Brabender, 1936). The error may be avoided by passing streams of air at definite humidities over the surfaces at such rates that no vapour-pressure gradients are formed.* This method (Wosnessensky and Dubnikow, 1936) was successfully developed.

The apparatus (figure 2) consists essentially of a humidifying train producing

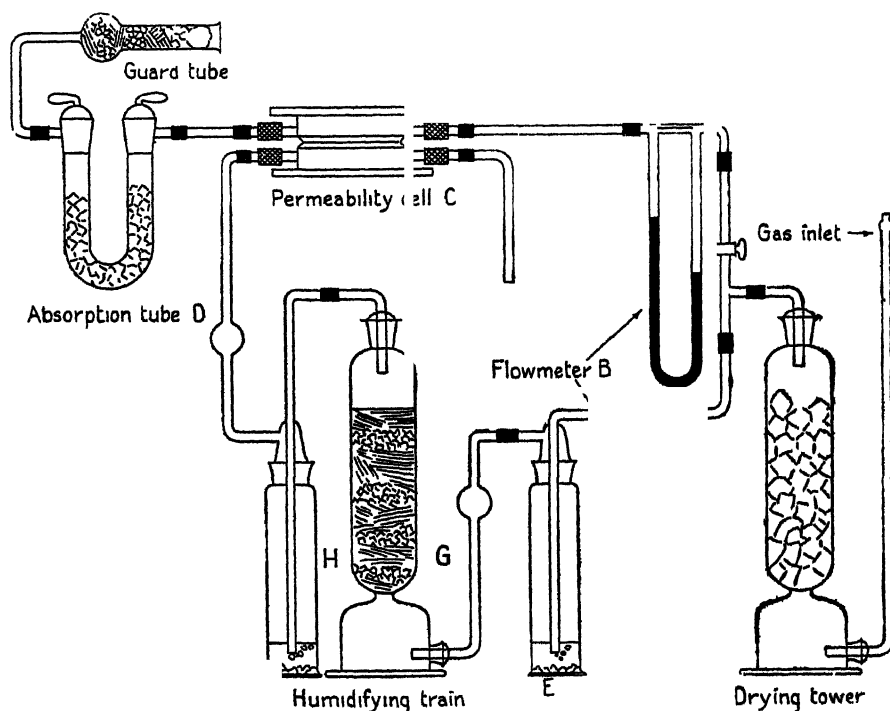


Figure 2 Moisture permeability apparatus

a stream of gas at the required humidity which then passes over the film, contained in the permeability cell. The moisture which permeates through the film is taken up by a dry-gas stream and then removed by drying tubes to determine the amount.

Gas, from a cylinder, is initially dried with silica gel and split into two streams, the rates of which are measured by the flow-meters B. One stream passes through the cell C to remove the water which has penetrated, and which is estimated gravimetrically by the absorption tubes of phosphorus pentoxide D. The other stream is pre-humidified by the water bubbler E. The vapour

* The authors are indebted to Dr. G. Barr for advice on this subject.

pressure is finally corrected by passing the gas through the tower of moist crystals G and the final bubbler H. After flowing over the film, the stream is released into the atmosphere.

It was found that a streaming rate of 1 litre per hour was sufficient to prevent the formation of any vapour-pressure gradient over the surfaces of the film and the flowmeters were constructed accordingly. The pressure difference developed across a capillary was measured, and the meters were calibrated by displacement of water, the necessary corrections being made.

To ensure further that the practical conditions correspond to those required the permeability cell was constructed with the following necessities in mind:—

1. A definite area of film must be exposed to the humid air or gas.
2. The only leakage from the moist to the dry side of the film must be through the film.
3. There must be no leakage from either side of the film to the outer atmosphere.
4. The thickness of the gas layers adjacent to the surfaces of the film must be small enough to prevent the formation of vapour pressure gradients.

The final form of permeability cell adopted and used for the experiments is shown in figures 3 and 4. The film of varnish was clamped between two shallow

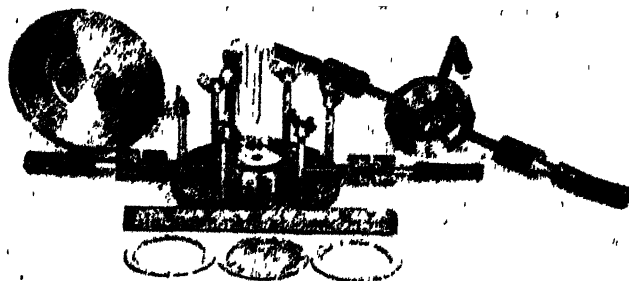


Figure 3. Moisture penetration cell. (General view.)

brass cylinders of internal diameter 1.5 in., which were held symmetrically in place by two steel plates provided with bolts and wing nuts. To ensure complete sealing, the areas of contact between the film and the cylinders were covered with rubber gaskets. Each compartment of the cell was provided with diametrically opposite openings to serve as inlet and outlet respectively for the gas streams and a third opening for connection to a small external manometer for checking the pressure and indicating leakages in the system.

The amount of moisture permeating through the films used in these experiments was of the order of 1 mg. per hour. To obtain five values of the permeation per hour within three days and with 5% accuracy the minimum weight measure was 3 mg., an analytical balance being used.

The use of phosphorus pentoxide supported on silica gel was found to be superior to that of any other absorbent and it was the most convenient to handle. Other absorbents used were silica gel, phosphorus pentoxide on glass wool and on pumice, and sulphuric acid on pumice.

a practically infinite volume of inert gas at the required humidity, or by measurement of the equilibrium vapour pressure in a finite volume initially containing a known quantity of the vapour. The latter method requires elaborate apparatus and technique which to a certain extent are avoided by use of the former.

Measurement of the mass sorbed, although simple in principle, presents some difficulties because of the small quantities involved. The most satisfactory weighing element is a quartz helix, which obtains the greatest sensitivity with a large ratio of length to cross-sectional area. Highly sensitive elements are, however, subject to vibration which impairs accurate observation. With the available instruments for measuring the extension of the helix, the limit sensitivity was found to be 4×10^{-5} g. for a load of 0.1 g., representing a percentage of 0.04 ± 0.02 . Beyond this limit, vibration frequently prevented accurate observation.

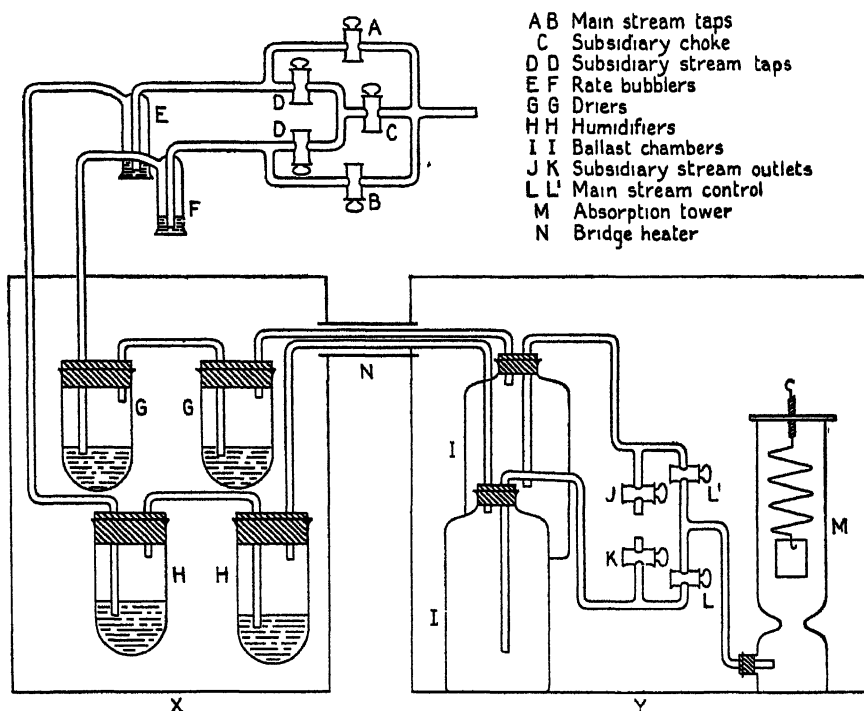


Figure 5. Apparatus used for determination of moisture sorption by films.

From the equations (3) and (4) it is evident that measurements of two kinds are required, of equilibrium absorption and of the rate of sorption. It is convenient to combine these in one experiment. Although it is not likely that the amount of sorption will appreciably alter the quantity of water vapour in the carrier gas, it is necessary, in experiments on the rate of sorption, that the humidity should change rapidly from the initial value to that at which the rate is being determined. Preliminary results showed that the sorption of films of the thickness used is practically complete within 30 minutes; consequently the time of change from the initial to the final state must not be greater than $\frac{1}{2}$ min. for

reasonable accuracy. Any method which depends on the mixing of static volumes of dry and humid air to give the required vapour pressure is thus uncertain, and a streaming method is preferable.

The final set-up is shown diagrammatically in figure 5, although most of the results were obtained with a simpler apparatus. The later arrangement has independent control of the temperature of the humidifying train, and consequently of the vapour pressure, and also of the film. The incoming air stream is split along three branches: A and B are the dry and humid streams respectively, and C is a subsidiary. The function of the last is to maintain the section of the apparatus not in use, comprising bubblers and ballast chambers, in either dry or humid condition, as the case may be, so that the change from one state to the other can be made in the shortest possible time. C acts as a choke, having a constant setting, and controls the rate of the slow maintaining stream, the direction being determined by the taps D, D'. It is released into the atmosphere by either J or K and prevented from entering the sorption tower by the taps L, L'. The humidifiers H, H' are filled with saturated salt solution and the driers G, G' with strong sulphuric acid.

Dry air was passed at a slow rate through the apparatus for a few hours to remove moisture from internal surfaces and from the film. The position of a mark on the bottom of the helix was read by a cathetometer external to the apparatus, illumination being supplied by a neon lamp behind the sorption tower. It was found that this reduces both aberration and eye-strain. The change to the humid stream was made by opening L and closing K, whereby the whole of the air passes through the humidifiers. The dry side of the apparatus was then isolated by closing the tap L'. Readings of the spring extension were taken initially at 1 min. intervals until the rate of sorption became slow, when the time was increased gradually. When the extension showed no positive change after $\frac{1}{2}$ hour, it was considered that the film was in equilibrium at the particular vapour pressure.

The rate of desorption was found in a similar manner, the operations being reversed.

§ 5. EXPERIMENTAL RESULTS

Permeability constants.

The permeability constants are calculated from equation (4), the unit being g.cm./hour cm² mm. Hg. In general the values for permeability constants of varnishes and similar materials are of the order of 10^{-8} . All values given are $P \times 10^8$.

The accuracy of the results for a particular film is of the order of 5% but, as is evident from inspection of the values, films of the same thickness prepared at different times do not show the same result. Where the values of the permeability are small, as in the case of phenol formaldehyde resin, $P \approx 0.5 \times 10^{-8}$, the accuracy is of a lower order because of the minute quantities permeating. All films were used one week after preparation. The results of the tests are given in tables 4 to 7.

Table 4. The moisture permeability constants of films of varnish 1

Temperature of test (°c.)	Thickness (cm.)	Vapour pressure diff. (mm. Hg)	$P \times 10^8$
25 35	0.0042	20.9 36.1	4.9 5.8
25	0.0083	21.9	4.8
30 35 40	0.0102	28.5 38.0 50.3	6.0 6.7 7.1

Table 5. The moisture permeability constants of films of varnish 2

Temperature of test (°c.)	Thickness (cm.)	Vapour pressure diff. (mm. Hg)	$P \times 10^8$
30	0.0053	28.8	2.3
30 40	0.0054	29.0 51.3	2.4 2.5
30 40	0.0062	28.7 50.7	2.6 2.7
30	0.0082	28.6	2.8

Table 6. The moisture permeability constants of films of varnish 3

Temperate of test (°c.)	Thickness (cm.)	Vapour pressure diff. (mm. Hg)	$P \times 10^8$
30 A 40 49.5	0.0028	29.7 52.6 85.5	0.42 0.35 0.34
25 B 35	0.0028	21.7 38.0	0.55 0.46
30 40 49.5	0.0045	29.6 52.5 85.2	0.59 0.45 0.44

Table 7. The moisture permeability constants of films of two baking varnishes at 25° c. and 96% R.H.*

Varnish No.	Thickness (cm.)	Vapour pressure diff. (mm. Hg)	$P \times 10^8$
510	0.0058	21.0	4.3
	0.0150	21.9	5.4
	0.0184	22.2	4.3
602	0.0062	21.7	0.57
	0.0106	21.9	0.42

* In view of the interest shown by varnish manufacturers in the great difference of moisture resistance between varnishes Nos 510 (Tung oil) and 602 (Tung oil—Congo Copal ester—linseed oil), tests are being carried out on the sorption of the latter to determine whether or not its properties are so modified by the sorbed moisture as to make the permeability constant small. The results will be reported later.

The films A and B, although of the same thickness, were prepared at different times.

The values for all three varnishes were obtained at 96% relative humidity.

The tests on film No. 1360 were carried out at 70% relative humidity, but for the remainder of the films the value was 96%. The results of the tests are given in table 8.

Table 8. The moisture permeability constants of polystyrene films

Film No.	Temperature (°c.)	Thickness (cm.)	Vapour pressure diff. (mm. Hg)	$P \times 10^8$
1360	25	0.0025	13.7	3.7
	30		17.9	3.9
	40		29.1	3.7
	50		51.8	2.9
1358	25	0.0024	21.0	2.3
1358A	25	0.0069	21.3	3.0
1359	25	0.0122	21.6	3.1

Values for the hydrogen and carbon dioxide permeation were found for a film of varnish No. 1 and for polystyrene film No. 1360. The difference of pressure across the film is not corrected for the amount permeated, the quantity being too small. The gas issues at constant pressure, hence the actual pressure of the gas is the atmospheric pressure. The results of the tests are given in tables 9 and 10.

Table 9. The gas permeability constants for a film of varnish 1

Gas	Temperature of test (°C.)	Thickness (cm)	Pressure diff. (mm. Hg)	$P \times 10^8$
H ₂	30	0.0081	754	0.0042
	40		773	0.0051
	50		771	0.0059
CO ₂	30	0.0140	754	0.050
	40		773	0.085
	50		771	0.153

Table 10. The gas permeability constants for a film of polystyrene 1360

Gas	Temperature of test (°C.)	Thickness (cm)	Pressure diff. (mm Hg)	$P \times 10^8$
H ₂	25	0.0025	760	0.0063
	25*		742	0.0064
	30		758	0.0076
	40		753	0.0091
	50		756	0.0104
CO ₂	25	0.0025	759	0.066
	25*		742	0.082
	30		758	0.069
	40		753	0.091
	50		758	0.106

* In this test both streams of gas contained water vapour at 70% relative humidity

Sorption constants

The sorption constant as defined by equation (6) is evidently the concentration in the film for a unit pressure of the gas or vapour sorbed. The unit in which the results are expressed is g./cm.³ per mm. of Hg.

These values are given in table 11 with the percentage sorptions by weight

Table 11. Moisture sorption constants for varnishes 1, 2 and 3

Varnish No	Thickness (cm) *	Sorption (%)	Sorption constant $\times 10^4$
1	A 0.0029	2.0	9.9
	B 0.0054	2.2	10.3
	C 0.0070	1.2	5.8
2	A 0.0074	1.0	5.1
	B 0.0086	0.9	4.4
3	A 0.0017	2.6	14.2
	B 0.0034	2.6	14.0
	C 0.0049	2.0	10.1

The letters A, B, C refer to different films.

for varnishes 1, 2, 3. All the values recorded were determined at 25° c. and 96% relative humidity.

The results are not all of equal accuracy. Films 1C, 2B and 3A were the first to be tested, whilst the method was being developed, and may be more in error. The order of accuracy is 7%.

Table 12. Moisture sorption constants for polystyrene films

Film No.	Thickness (cm.)	Vapour pressure (mm. Hg)	Sorption (%)	Sorption constant $\times 10^4$
1358	0.0024	17.9	0.10	0.6
1358A	0.0069	17.9	0.08	0.4
1359	0.0122	17.9	0.07	0.3
		21.4	0.11	0.4

Because of the extremely small values of the sorption of the polystyrene films with consequent uncertainty of the measurements, six determinations were made for each film and the mean taken. The accuracy of the sorption percentages is ± 0.02 , with a percentage accuracy of *c.* 20.

Rate of sorption

The rates of sorption are not reported numerically, requiring too much space, but specimens of the curves are given in figures 6, 7, and 8 for the three varnishes.

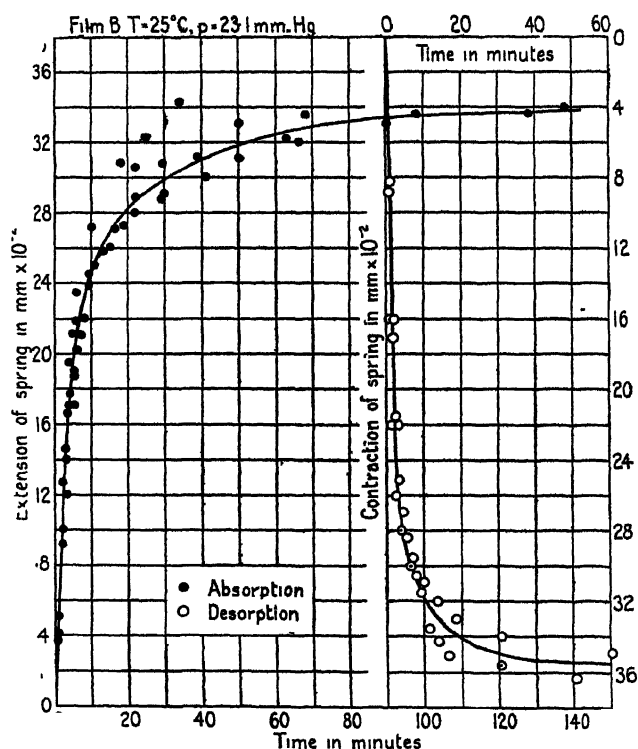


Figure 6 Sorption vs time for films of tung oil-phenol formaldehyde varnish No. 1.

The points are the results of three rate determinations for both sorption and

Using the relation developed in equation (11), the values of $\log (m_s - m)$ are plotted against t in figures 9, 10 and 11.

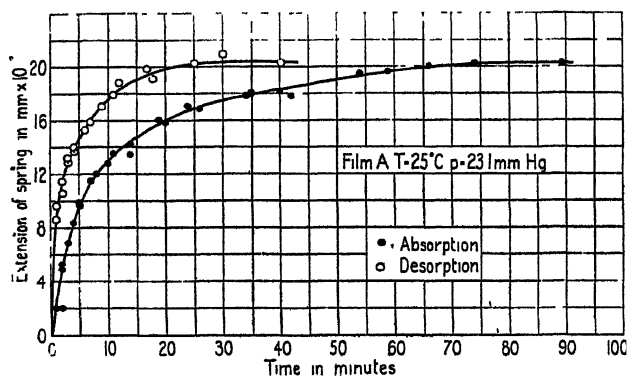


Figure 7. Sorption vs time for films of linseed oil-gilsonite varnish No. 2.

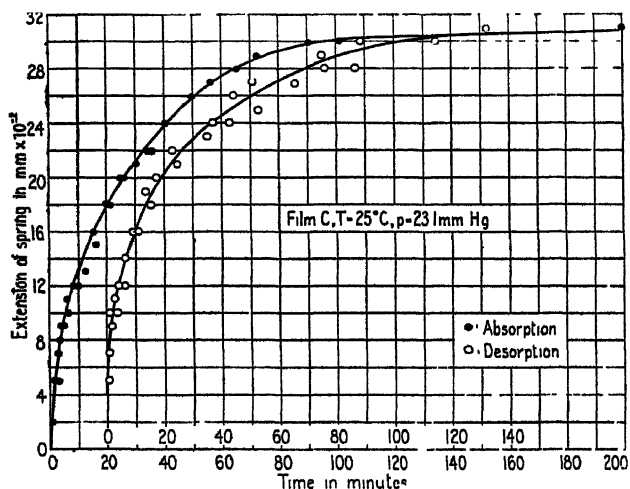


Figure 8 Sorption vs. time for films of phenol formaldehyde resin varnish No. 3

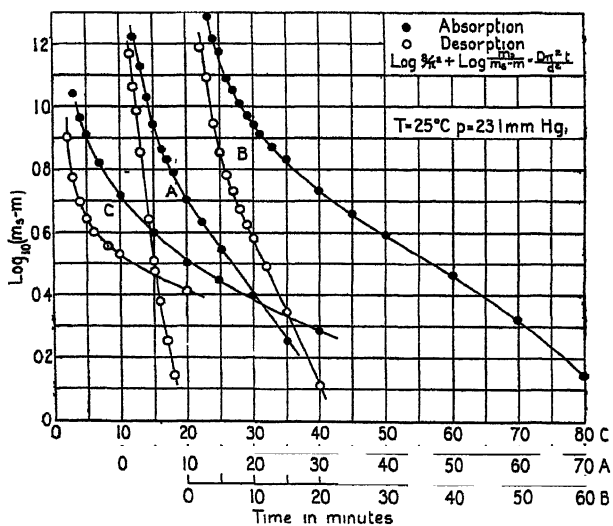


Figure 9 Log residual sorption vs. time for films of tung oil-phenol formaldehyde

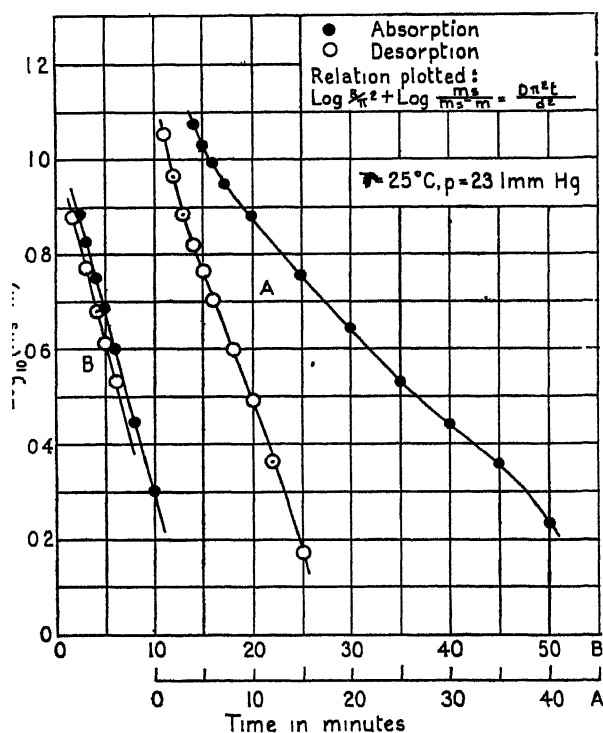


Figure 10 Log residual-sorption vs time for films of linseed oil-gilsonite varnish No. 2.

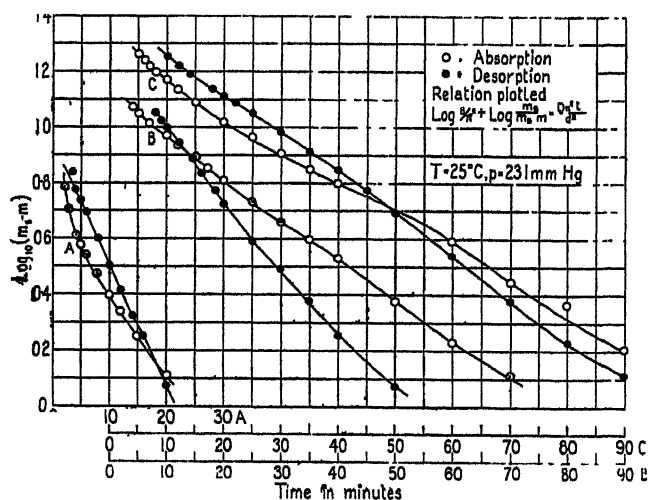


Figure 11 Log residual-sorption vs. time for films of phenol formaldehyde resin varnish No. 3.

§ 6. DISCUSSION

Nature of the permeation mechanisms

The formal theory already developed (§ 2) assumed that diffusion in solid materials takes place by a process analogous to osmotic diffusion in liquids and not by flow through micro-pores. Such a hydrodynamic flow would be expected

to lead to a variation of P with $1/\sqrt{T}$, where T is the absolute temperature. The results of the permeation of gases through varnish 1 and polystyrene show that the temperature coefficient is much too large. Simplifying the relationship to

$$P = k/\sqrt{T}, \quad \dots \dots (12)$$

and using the values of P from table 10 for hydrogen permeation through polystyrene, the values of k , which should be a constant, vary from 10.8 to 18.9 as T varies from 25° c. to 50° c. Further, the temperature coefficient of P , which should be negative according to equation (12), is usually positive.

If, therefore, permeation of gases does not occur as flow through micro-pores in the materials concerned, it can be concluded that the same holds for the permeation of water and other substances.

A crucial test of the theory of an osmotic diffusion of water through varnishes and similar materials is the comparison of the permeability constant determined in the stationary state and that calculated by equation (9) from the diffusion constant obtained from measurements in the non-stationary state using equation (11). In this comparison all the assumptions are involved. Inspection of figures 9, 10 and 11, which show the relation between $\log(m_s - m)$ and t , indicates that equation (11) is reasonably valid. From the curves, values of the diffusion constant D have been derived, the unit being $\text{cm}^2/\text{hr.}$ (Table 13.)

Table 13. Diffusion constants for varnishes 1, 2, 3 at 25° c.

Varnish No.	Thickness (cm.)	Mean value of $D \times 10^8$ ($\text{cm}^2/\text{hr.}$)
1	0.0029	12.5
	0.0054	27.0
2	0.0074	34.5
	0.0086	82.5
3	0.0017	2.0
	0.0034	3.2
	0.0049	5.0

These values are the means for the sorption and desorption processes.

From these values, and using the sorption constants of table 11, the permeability constants (table 14) have been calculated from equation (9) and compared with those determined experimentally and given in tables 4, 5 and 6.

The agreement is good in showing the correct order of magnitude. Considering also the approximate constancy of the permeability constant with thickness, there is reason to believe that the mechanism of permeation is one of osmotic diffusion, and that deviations from the formal laws are to be explained in terms of the characteristics of the substance permeating and the material permeated.

Table 14. Experiments and calculated values for the permeability constants at 25° C.

Varnish No.	Thickness (cm.)	Permeability constant $\times 10^*$ (g /hr cm.mm Hg)	
		Direct observation	Calculated from D and S
1	0 0029	4.8*	1 2
	0 0054	4 9	2 8
2	0 0074	2 6*	1 8
	0 0086	2.7*	3 6
3	0 0017	0 42*	0 28
	0.0034	0 45*	0.44
	0 0049	0 63*	0 51

* The values are interpolated

Deviation from ideal behaviour

Examination of tables 4, 5 and 6 shows that permeability constants tend to be larger for thicker films, a result also noted by other authors (Edwards and Wray, 1936; Lishmund and Siddle, 1941, Wray and van Vorst, 1941). The increasing sorption of polystyrene with decreasing thickness (table 12) indicates the presence of a concentration on the surface of the film over and above that existing inside the material. Calculation shows that this surface layer is equivalent to a thickness of 4 molecules. Such a layer may have a lower permeability than the interior, and thus give rise to the variation with thickness mentioned.

The effect of vapour pressure on the permeability constant has not been determined in the present investigation, but investigations on baking varnish (Edwards and Wray, 1936), rubber (Lowry and Kohmann, 1927; Taylor, Hermann and Kemp, 1936), and cellophane (Simril and Smith, 1942) show that, in general, there is a considerable increase of sorption with vapour pressure, especially in the region of 80–100% relative humidity. This super-sorption must be a consequence of the polar nature of the water molecule as opposed to the non-polar character of a gas, e.g. hydrogen, the solubility of which closely follows the linear law. First there may be a liquid condensation on the hydrophilic groups in the material, and secondly, intra-micellar penetration of water molecules, which causes swelling and exposure of more hydrophilic groups. Payne (1939) has shown that the permeability of a material is greatest with respect to a vapour which in its liquid state dissolves or disperses the material. If liquid sorption takes place, it is to be expected that the binding force at a high concentration will be smaller than that exerted when the water molecule is sorbed directly on a hydrophilic group. Thus movement at higher concentrations will be freer, and if there is some disruption there will be more traversable paths.

Variation with temperature

All the considerations so far have dealt with processes at constant temperature. It was stated in § 6 that the variation of the gas permeability with temperature is not only in the wrong direction, but of too great a magnitude to be accounted for by a flow through micro-pores. Barrer (1941) showed that permeation and diffusion of gases through organic materials was by an activated mechanism, similar to that of the movement of atoms and ions in fluid and solid media. It was found that the variation of permeability and diffusion constants with temperature could be represented by the equations

$$P = P_0 \times \exp(-E/RT), \quad \dots\dots(13)$$

$$D = D_0 \times \exp(-E'/RT), \quad \dots\dots(14)$$

where the E and E' are activation energies. The activation energy for the permeation process is compounded of that for diffusion and that for sorption, this latter being in fact the heat of solution.

It has been calculated from equation (13) that the activation energy for the permeation of hydrogen and carbon dioxide through polystyrene (table 10) is 3800 calories per gramme molecule in both cases, and for the permeation through varnish 1 (table 9) the energies are 3400 calories per gramme molecule and 11,400 calories per gramme molecule for hydrogen and carbon dioxide respectively. The first two values are lower than those found by Barrer for permeation of hydrogen through rubbers, ebonite, bakelite and cellulose derivatives, the range being 5000 to 10,000 calories per gramme molecule.

Although the permeability constants of the varnishes, given in tables 4 and 5, increase with temperature, the fact that the vapour pressure also increased does not allow a decision to be made. A similar consideration applies to the interpretation of the negative temperature coefficient of P for phenol formaldehyde resin and the constancy of P for polystyrene. Until the variations of permeability constant with temperature and vapour pressure are studied independently, it will not be possible to decide whether or not the process is activated. It is clearly of the greatest importance to determine this since Barrer has shown that the activation energies for gas permeation are closely related to the energies associated with mechanical changes in the case of rubbers.

Comparison of diffusion rates

It has been urged that diffusion of water through varnishes and similar materials is of a fundamentally different nature from that of gases in view of the great differences in the permeability constants for the two substances. These differences are magnified in two ways, first by the fact that the permeability constant is in terms of the mass permeating and not of the numbers of molecules, and secondly, because the sorption of water is much greater than that of gases. The first distortion is easily overcome by dividing the constants by the molecular weight of the gas or vapour permeating; it is then in terms of gramme molecules, and consequently directly proportional to the number of molecules passing. The order of permeability constant is then

$$\text{Water vapour} = 0.3 \times 10^8, \quad \text{gas} = 0.002 \times 10^8$$

for a film of varnish 1 at 30° C. The different effects of sorption are more difficult to eliminate, the true comparison being between the diffusion constants

which give the amounts moving through the material under unit concentration gradient within the material. The diffusion constant is thus independent of the degree of sorption (see Gettens and Bigelow, 1933). Only a few values of D are known, and these have been collected in table 15

Table 15. Moisture diffusion constants for several materials

Substance	°C.	$D \times 10^6$ (cm ² /sec.)
Varnish 1	25	0.013
Varnish 2	25	0.015
Varnish 3	25	0.0011
Polystyrene	25	0.21
Values quoted by Ferry (1941)		
Polyvinyl acetal	20	0.047
	40	0.062
Polyvinyl butyral	25	0.0035
Polyvinyl acetate	40	0.0056
Methyl methacrylate	25	0.027
Phenol formaldehyde	25	0.0022

There are no figures for the diffusion constant for gases permeating through any of the materials of table 15. The order of the value of the diffusion constant may, however, be obtained by assuming the average value for the solubility of hydrogen in organic materials; it is generally in the region of 0.05 ml. at N.T.P./cm³ of solid/atm., from which the sorption constant may be found. From the known values for the permeability constant the diffusion constants may be calculated, and these are given in table 16.

Table 16. Calculated hydrogen diffusion constants at 25° c.

Substance	°C.	$D \times 10^6$ (cm ² /sec.)
Varnish 1	25	1.7
Varnish 3	25	0.017
Polystyrene	25	3.0

In all cases the diffusion constant is greater for the gas than for moisture (see table 15), the variation being greater than any probable deviation of the hydrogen solubility from the value assumed. It is probable the solubility will be smaller, which will increase the diffusion constant. The values for varnish 1 and polystyrene are of the order of those which Barrer found in 1941 for rubbers; the lower value for phenol formaldehyde is in accord with the low value of the moisture diffusion constant (table 15) and probably results from its structure.

Structure and diffusion

The effect of the external factors, vapour pressure and temperature, has been considered on the assumption that the hydrophilic groups are of the greatest importance in determining the sorption and permeation characteristics. Whilst,

however, a satisfactory explanation of the deviations and variations of the constants is possible, the factors determining the absolute diffusion rate and sorptive capacity depend on the structure of the material. It is to be expected that both the chemical nature of the ultimate molecular units and the manner in which these are aggregated into the solid will be the determining factors.

Practically all the materials used as organic insulating materials consist of giant molecules whose structure is a repetition of simple units. Such a repetition may be chain-like as in substances such as polystyrene, tung oil, rubber, or it may be three-dimensional, as in vulcanized rubber, and the condensation products of formaldehyde with phenol, urea or melamine. In the chain-polymeric materials, molecular aggregates, termed micelles, are constituted by secondary valence forces, but in the three-dimensional polymers a more rigid structure arises by primary valence forces cross-linking the molecules. However, the energies necessary to break a primary and a secondary valence bond are about 30,000 and 6000 calories per Avogadro number of bonds broken respectively, so that the probability of a secondary valence bond rupturing is

$$[\exp(-6000/RT)/\exp(-30,000/RT)] = 10^{20} \text{ times}$$

more than that of a primary bond. The values of table 16 show that the difference of permeation cannot be accounted for by such a difference in mechanism. In fact, no activation energies for permeation of gases through various membranes are much above 10,000 calories per gramme molecule. The conclusion is, therefore, that diffusion takes place along the micelle boundaries and through holes arising from the rupturing of secondary valence bonds, and differs only in degree in chain-like and three-dimensional polymers. A unifying concept is that of a structure consisting of regions of order and disorder, diffusion occurring through the disordered regions at a rate depending on their volume relative to the whole. The activation energy is then a measure of the energy necessary to produce sufficient disorder in the ordered regions for the molecule to pass through, and its magnitude depends on the cohesive forces between the molecules or parts of molecules in the ordered region. On this basis the low diffusion rate through paraffin wax (which has an ordered molecular lattice, providing no inter-crystalline cracking is present) may be explained as well as the decreased permeability of polystyrene film when it is partly oriented by stretching (Badam and Leilich, 1938). Further, the increased order produced by cross-linkages in three-dimensional polymers will lower the rate of diffusion, as is shown by Sager (1934).

§7 CONCLUSIONS

The experimental methods which have been developed are capable of giving consistent results to an accuracy of about 5% for the permeability measurements and about 7% for the sorption measurements. The most likely source of error in both cases results from the use of saturated salt solutions for humidification of the gas streams. For the varnishes a further error is introduced in that similar films prepared from the same material show permeability characteristics which may vary by 4%. Such variation is likely to arise from inequality of thickness. At the present stage of the research, which is largely exploratory, these errors are not serious, and the methods, capable of improvement if necessary, provide adequate means of investigation.

The results of this and other investigations show that the moisture permeability constants for varnishes and lacquers have a 10-fold range, which is about in the middle of the 10^4 -fold range of constants for all classes of dielectrics. There is, in general, an increase of permeability constant with increase of vapour pressure and temperature, although the extent to which either of these separately is responsible has not yet been determined. The degree of variation of the constant varies widely.

A variation of permeability constant with thickness of film also occurs, but this may be due to a slow surface process.

The relationship between permeability, diffusion and sorption constants approaches that predicted by theory. The deviations arise because the assumptions on which the theory is based take no account of the real nature of sorption phenomena nor of the molecular structure of the material. It is assumed that during the process of permeation no physical interaction, other than simple solution, takes place. Actually this is not so, and to interpret the facts of permeability and sorption the concepts of molecular physics must be employed.

It is not yet possible to give a detailed picture of the mechanism of permeability. Nevertheless, some observations on the effect of structure are permissible:

1. The greater the occurrence of water-attractive groups the higher the permeability of the substances.
2. The greater the degree of order in the structure, that is, the greater the regular molecular orientation, the less is the permeability.

These two factors acting together determine the permeability of any organic insulating material of the types considered. In varnishes the effects are about balanced; more permeable substances, such as cellulose derivatives, have much greater sorptive powers; the less permeable materials, such as highly vulcanized rubbers, and waxes, have more ordered structure. The importance of the structure of the material has been recognized in the manufacture of Styroflex oriented polystyrene, which has only half the moisture permeability of the randomly oriented material.

Further, the two factors are clearly important for the dielectric and mechanical behaviour of insulating materials. Thus it seems possible that relationships between these properties and the moisture permeability may be developed in terms of the physical and chemical properties of the molecules.

ACKNOWLEDGMENTS

The authors are indebted to Mr. N. A. Bennett for advice on the types of varnish on which the research should be carried out, to Messrs. Griffith Bros. (London), Ltd., who supplied the experimental varnishes referred to in table 1, and to the Director of the British Electrical and Allied Industries Research Association for permission to publish this paper.

APPENDIX

References to moisture permeability constants cited in the literature

Moisture permeability constants for a considerable range of materials have been determined by various workers, and collections of these values will be found in the publications of Taylor, Hermann and Kemp (1936), Carson (1937), and

Barrer (1941). The dispersion of these values is shown diagrammatically in figure 12.

Values of the moisture permeability constant are gathered together in table 17 for varnish and lacquer films, paints being omitted since the pigment affects the permeability. Except for the values of Lishmund and Siddle (1941), which were determined by a method somewhat similar to that described in this report, all of the permeability constants were found by the use of a variation of the Muckenfuss pot.

In essentials this apparatus consists of a cup or jar over the mouth of which the varnish film is fixed. The cup may contain a desiccant, in which case the gain in weight is found on exposure to a humid atmosphere, or a humidifying agent may be used in the cup and the loss in weight in a dry enclosure found. The desiccating or humidifying surface in the cup may be some way from the

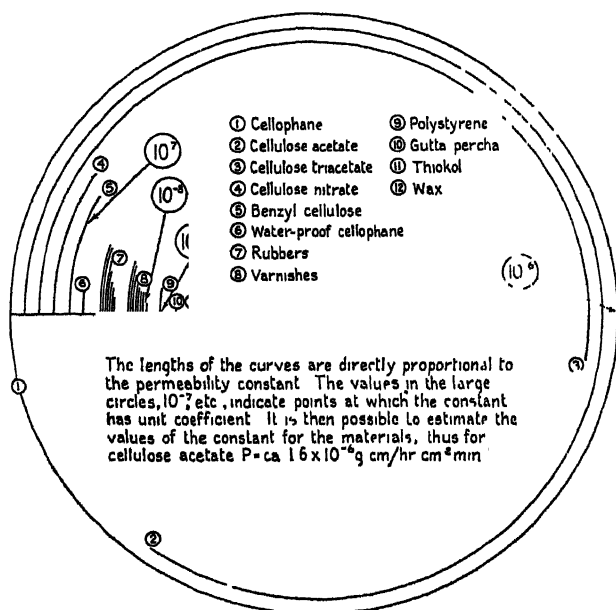


Figure 12. Moisture permeabilities of some non-porous insulating materials

surface of the film, resulting in a gradient of vapour pressure due to the finite time of diffusion of water vapour in air. It has been calculated that if the distance is not greater than 1 cm. no appreciable error is involved for permeability constants of the order of 10^{-8} . For larger values of the constant, as for cellulosic materials, $P = 10^{-6}$, the error is of the order of 70%. With elevated temperatures, difficulty is experienced in sealing the film to the cup: sealing at room temperature results in ballooning of the film on heating, due to expansion of the air. With these limitations the method is useful for industrial purposes.

The values for the permeability constant are given in the units used throughout this report and under the following conditions whenever possible:—

25° C., 96% R.H., 0.005 cm. thickness.

Values at other temperatures, vapour pressure differences or thicknesses are to be found in the references quoted.

Table 17. Moisture permeability constants for varnishes as free films at 25° C., 96% R.H. and a thickness of 0.005 cm.

Material	$P \times 10^8$ (g cm/hr cm ² mm.Hg)	Material	$P \times 10^8$ (g cm/hr. cm ² mm.Hg)
<i>Nitrocellulose</i>		<i>Synthetic</i>	
†alkyd resin (1)	0.4 3.0	Vinyl resin (1)	0.8
		Glycerol phthalate (4)	1.3
		„ „ (5)	6.2
†alkyd resin + dibutyl phthalate (2)	4.0	„ „ (4)	0.3
†alkyd resin + dibutyl phthalate + castor oil (2)	4.5	Phenol formaldehyde resin (4)	1.2
†dammar (2)	2.5	„ „ „ (6)	0.6
†dammar + dibutyl phthalate (2)	3.3	Alkyd resin (7)	7.2
†rosin-maleic acid + dibutyl phthalate (2)	1.1	<i>Oil varnishes</i>	
†ester gum + dibutyl phthalate (3)	1.7	Ester gum (4)	0.9
†diphenyl resin (1)	0.2	Tung oil + phenol resin (7)	2.1
†vinyl resin (1)	6.0	„ „ „ (3)	4.2
†phenolic resin (1)	0.9	„ „ „ (6)	4.9
†dibutyl phthalate (2)	5.3	Tung oil + linseed oil + phenolic resin (5)	5.1
		Linseed oil + alkyd resin (3)	5.0
		Linseed oil + bitumen (6)	2.3
		Orange shellac (7)	2.5

(1) Wray and Van Vorst, 1936.

(2) Wing, 1936.

(3) Lishmund and Siddle, 1941.

(4) Edwards and Wray, 1936.

(5) Kline, 1937.

(6) This paper, 1945.

(7) Payne, 1939.

REFERENCES

- ABRAMS and BRABENDER, 1936. *J. Paper Trade*, T.S.32, 102.
 BADUM, E. and LEILICH, K., 1938. *Felten u. Guilleaume Rdsch.* 22, 13.
 BARRER, R. M., 1941. *Diffusion in and through Solids* (Cambridge: The University Press).
 CARSON, F. T., 1937. *Nat. Bur. Stand. Misc. Publication* M127.
 EDWARDS, J. D. and WRAY, R. I., 1936. *Ind. Eng. Chem.* 28, 549.
 GETTENS, R. J., 1932. *Tech. Studies Field Fine Arts*, 1, 63.
 GETTENS, R. J. and BIGELOW, E., 1933. *Tech. Studies Field Fine Arts*, 2, 15.
 IRANY, P., 1941. *Ind. Eng. Chem.* 33, 1551.
 KLINE, G., 1937. *Bur. Stand. J. Res.*, Wash., 18, 235.
 LISHMUND, R. E. and SIDDLE, F. J., 1941. *J. Oil Col. Chem.* 24, 122.
 LOWRY, H. and KOHMANN, G., 1927. *J. Phys. Chem.* 31, 23.
 PAYNE, H. F., 1939. *Ind. Eng. Chem. (Anal.)*, 453, 11.
 PAYNE, H. F. and GARDNER, W. H., 1937. *Ind. Eng. Chem.* 20, 827.
 SAGER, T. P., 1934. *Bur. Stand. J. Res.*, Wash., 13, 879.
 SIMRIL and SMITH, 1942. *Ind. Eng. Chem.* 34, 226.
 TAYLOR, R. L., HERMANN, D. B. and KEMP, A. R., 1936. *Ind. Eng. Chem.* 28, 1255.
 WING, H. J., 1936. *Ind. Eng. Chem.* 28, 788.
 WOSNESSENSKY, S. and DUBNIKOW, L. M., 1936. *Kolloid-Z.* 74, 183.
 WRAY, R. I. and VAN VORST, A. R., 1936. *Ind. Eng. Chem.* 28, 1268.

A SYSTEM OF TRANSFER COEFFICIENTS FOR USE IN THE DESIGN OF LENS SYSTEMS: I. THE GENERAL THEORY OF THE TRANSFER COEFFICIENTS

By F. D. CRUICKSHANK,
University of Tasmania

MS. received 25 January 1944, in revised form 26 February 1945

ABSTRACT A detailed account is given of a series of differential transfer coefficients which specify the changes in the path of a ray in the final image-space resulting from small changes in surface curvature, refractive index and axial separation of successive surfaces, made within the system. Extending the method to the intersection points of the typical rays of a standard trigonometrical trace, transfer coefficients are derived which specify the rate of change of the tangential aberrations with curvature, refractive index and axial separation at each surface of the system. In particular, coefficients are developed which relate the chromatic aberrations in a very simple manner to the dispersions of the glasses of the system.

§1 INTRODUCTION

M'AULAY and CRUICKSHANK (1945) have described a simple means of specifying the change in the path of a ray through a lens system resulting from a small alteration made at some surface of the system. Using the original traced path of the ray as a reference frame, the new path is completely defined when the values of the differential quantities dp' , dU' are known for the ray at each surface. Particular interest attaches to the values of these quantities at the last surface of the system, for from these the new path of the ray in the final image-space may be deduced. The basic problem is to relate the change in the path of the ray in the final image-space to the alteration within the system which produced it. A solution of this problem is presented in this paper in terms of a system of transfer coefficients, the values of which can be computed for any type of alteration made at any surface of the system. Once a system of transfer coefficients is established, the way is open for the development of designing methods of considerable power. In the paper referred to above, it is shown how the existence of the transfer coefficients leads to a systematic method of making the final correction of the aberrations of a lens system. In subsequent papers it will be shown how an analysis of the surface contributions to the aberrations of finite pencils may be made, and a direct method will be proposed for the estimation of the tolerances to be observed in the production of an optical system. The notation employed follows that of Conrady (1929).

§2 THE TRANSFER COEFFICIENTS FOR dU' -CHANGES AND dp -CHANGES

Suppose that the path of a selected ray has been traced trigonometrically through a system of k spherical refracting surfaces. After refraction at surface i of the system, the path of the ray is described by the quantities L_i' , U_i' . Let a

small change be made now at surface i such that the ray after refraction at the same incidence point on the surface takes the new direction specified by $U_i' + dU_i'$. The path of the ray through the remainder of the system will be different from the original traced path, and the ray will finally emerge from the last surface at a point displaced by some amount, dp_k' , from the emergence point of the traced ray, and in a direction inclined at an angle, dU_k' , to that of the traced ray. We may express these differential quantities, dp_k' , dU_k' , in terms of dU_i' by writing

$$dU_k = \frac{\partial U_k'}{\partial U_i'} dU_i', \quad \dots (1)$$

$$dp_k' = \frac{\partial p_k'}{\partial U_i'} dU_i'. \quad (2)$$

Similarly, if a small change made at surface i results only in an incidence point displacement, dp_i , the inclination angle, U_i , of the ray remaining unchanged, the ray will emerge from the last surface of the system at a point and in a direction specified by some other values, dp_k' , dU_k' , relative to the emergent traced ray. For these we may write

$$dp_k' = \frac{\partial p_k'}{\partial p_i} dp_i, \quad \dots (3)$$

$$dU_k = \frac{\partial U_k'}{\partial p_i} dp_i. \quad (4)$$

Equations (1) to (4) define the four *transfer coefficients*, $\partial p_k' / \partial U_i'$, etc., which can be calculated for a traced ray at each surface of the system. The method of calculation depends upon relations deduced in the next section. The transfer coefficients specify the effect of a dU' -change or a dp -change made at any surface on the path of the ray as it leaves the last surface of the system. These particular quantities are chosen as the fundamental transfer coefficients because any real change made in the course of an actual design can be described in terms of either a single dU' -change or a single dp -change.

§3 THE RELATION BETWEEN THE TRANSFER COEFFICIENTS AT SUCCESSIVE SURFACES OF THE SYSTEM

Denoting any two successive surfaces of the system by the subscripts i and j , we follow out the effect on the ray-path of a change dU_i' made at surface i . From

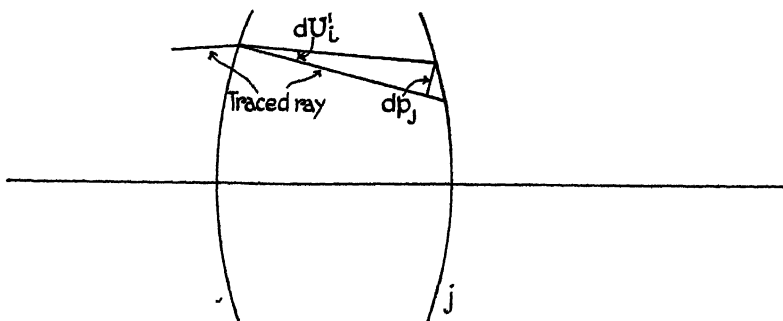


Figure 1

figure 1 it is seen that, relative to the traced path, the ray will now meet surface j with an incidence-point displacement $dp_j = -D_i' dU_i' = D_j dU_i'$, and with an

inclination angle change $dU_j = dU'_j$, where $D_j = -D'_j$ is the distance between the two surfaces measured along the traced ray. Further, the change dU_j alone would produce in turn a change $dU'_j = (\partial U'_j / \partial U_j) dU_j$ in the direction of the ray after refraction at surface j . Hence, the change dU'_i at surface i is equivalent in effect to the change $dp_i = D_i dU'_i$, together with the change $dU'_j = (\partial U'_j / \partial U_j) dU_j$, at surface j . Expressing the resulting dp'_k , dU'_k in terms of each of these equivalent changes, we have

$$\begin{aligned} \frac{\partial U_k}{\partial U'_i} dU'_i &= dU'_k = \frac{\partial U_k}{\partial U'_j} dU'_j + \frac{\partial U_k}{\partial p_j} dp_j \\ &= \frac{\partial U_k}{\partial U'_i} \frac{\partial U'_j}{\partial U_j} dU'_i + \frac{\partial U_k}{\partial p_j} D_j dU'_i, \end{aligned}$$

whence
$$\frac{\partial U_k}{\partial U'_i} \left(\frac{\partial U'_j}{\partial U_j} \frac{\partial U'_j}{\partial U_j} + \frac{\partial U_k}{\partial p_j} D_j \right), \quad \dots\dots(5)$$

and similarly,
$$\frac{\partial p'_k}{\partial U'_i} = \left(\frac{\partial p'_k}{\partial U_j} \frac{\partial U'_j}{\partial U_j} + \frac{\partial p'_k}{\partial p_j} D_j \right). \quad \dots\dots(6)$$

Next we follow out the effect on the ray path of a small change of incidence point, dp_i , made at surface i . It is seen from figure 2 that relative to the traced

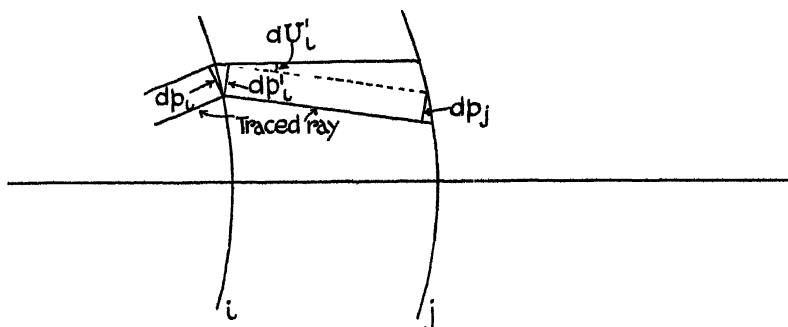


Figure 2

path the ray will now leave surface i after refraction with a change in inclination angle $dU'_i = (\partial U'_i / \partial p_i) dp_i$ and with an incidence-point displacement

$$dp'_i = (\partial p'_i / \partial p_i) dp_i,$$

the latter having the same effect as an equal displacement, dp_j , at surface j . Expressing the resulting dU'_k , dp'_k in terms of dp_i and the equivalent set of changes dU'_i , dp'_i , leads to the relations

$$\frac{\partial U_k}{\partial p_i} = \frac{\partial U_k}{\partial U'_i} \frac{\partial U'_i}{\partial p_i} + \frac{\partial U_k}{\partial p_j} \frac{\partial p'_i}{\partial p_i}, \quad (7)$$

$$\frac{\partial p'_k}{\partial p_i} = \frac{\partial p'_k}{\partial U'_i} \frac{\partial U'_i}{\partial p_i} + \frac{\partial p'_k}{\partial p_j} \frac{\partial p'_i}{\partial p_i}. \quad (8)$$

The four equations (5) to (8) permit the calculation of the transfer coefficients for surface i when those for surface j are known. At the last surface of the system it is obvious that the transfer coefficients have simple values, for

$$\frac{\partial U_k}{\partial U'_k} = 1 \quad \text{and} \quad \frac{\partial p'_k}{\partial U'_k} = 0,$$

while $\partial U'_k / \partial p_k$ and $\partial p'_k / \partial p_k$ are calculated in the usual way (see M'Aulay and Cruickshank (1945), equations (2) and (8)). The transfer coefficients for the surface $(k-1)$ may now be obtained by equations (5) to (8), and so the computation may be continued surface by surface through the system until the first surface is reached. At each surface the computation involves four pairs of multiplications, there being a common factor in each pair, and four additions.

§4 THE RAY TRACE AND THE TRANSFER COEFFICIENTS FOR THE INTERSECTION POINTS

To investigate the spherical aberration of a system it is usual to trace three rays from an axial object point, the marginal, zonal and paraxial rays. These rays have intersection points with the principal axis in the final image space denoted respectively by $M(L'_m, 0)$, $Z(L'_z, 0)$, and $P_x(l', 0)$, their positions being specified by rectangular coordinates relative to the pole of the last surface. For each oblique pencil traced for the extra-axial aberrations, three rays are selected, denoted by the symbols a , pr and b in the Conrady notation. Associated with these rays are the points $Ab(L'_{ab}, H'_{ab})$, the intersection of the rays a and b , $Pr(L'_{pr}, H'_{pr})$, the point at which the principal ray cuts the plane through Ab at right angles to the axis; $Prf(l', H'_{prf})$, the point at which the principal ray strikes the plane through the paraxial focus at right angles to the axis. In addition, we shall be concerned with the ideal image point, $Id(l', H'_{id})$, which shows where the principal ray would cut the paraxial image plane if the paraxial value of the magnification held throughout the field. If a change is made in the path of any of these typical rays at any surface of the system, the associated intersection point or points in the image space will be displaced. We now develop transfer coefficients to specify such changes.

The axial intersection points, M , Z and P_x . We consider first the effect of the change, dU'_i , on the position of the point M . Let us suppose that before the change is made in the ray path at surface i the marginal ray leaves the last surface of the system along the path PM , while after the change is made it follows the path $P'M'$ in figure 3. Relative to the traced ray PM , the displacement of

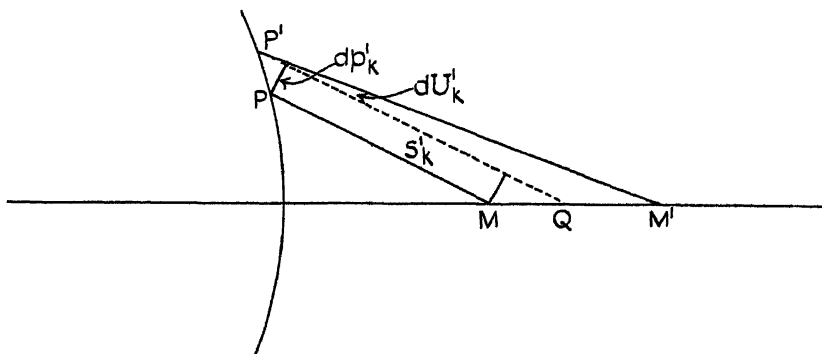


Figure 3

the emergence point is $dp'_k = (\partial p'_k / \partial U'_i) dU'_i$, and the change in direction is $dU'_k = (\partial U'_k / \partial U'_i) dU'_i$, and each of these contributes to the shift of M . The effect of the displacement dp'_k alone would be to move the intersection point

from M to Q, while the direction change dU_k' would, in addition, move it from Q to M'. For differential changes we have, to first order,

$$MQ = dp_k' \operatorname{cosec} U_{mk}', \quad QM' = -S_{mk}' dU_k' \operatorname{cosec} U_{mk}',$$

where S_{mk}' denotes the distance PM. Hence we have,

$$\frac{\partial L_m'}{\partial U_i'} dU_i' = MM' = \left(\frac{\partial p_k'}{\partial U_i'} dU_i' - S_k' \frac{\partial U_k'}{\partial U_i'} dU_i' \right)_m \operatorname{cosec} U_{mk}',$$

$$\text{whence} \quad \frac{\partial L_m'}{\partial U_i'} = \left(\frac{\partial p_k'}{\partial U_i'} - S_k' \frac{\partial U_k'}{\partial U_i'} \right)_m \operatorname{cosec} U_{mk}'. \quad \dots\dots (9)$$

Similarly, the consideration of the effect of a change, dp_i , leads to

$$\frac{\partial L_m'}{\partial p_i} = \left(\frac{\partial p_k'}{\partial p_i} - S_k' \frac{\partial U_k'}{\partial p_i} \right)_m \operatorname{cosec} U_{mk}'. \quad \dots\dots (10)$$

It is convenient to introduce the symbol C as an abbreviation, its meaning being defined by

$$C(U_i') = \left(\frac{\partial p_k'}{\partial U_i'} - S_k' \frac{\partial U_k'}{\partial U_i'} \right), \quad \dots\dots (11)$$

$$C(p_i) = \left(\frac{\partial p_k'}{\partial p_i} - S_k' \frac{\partial U_k'}{\partial p_i} \right). \quad \dots\dots (12)$$

Quantities of this C -type enter characteristically into all expressions for the transfer coefficients for the intersection points, as we shall see presently. Actually, the C -quantity measures the rate of change of the intersection point with the variable concerned along a direction at right angles to the traced ray. Where it is necessary to distinguish the C -quantities of different rays we do so by writing the appropriate suffix, e.g. $C(p_i)_m$. Rewriting equations (9) and (10)

$$\frac{\partial L_m'}{\partial U_i'} = C(U_i')_m \operatorname{cosec} U_{mk}', \quad \dots\dots (13)$$

$$\frac{\partial L_m'}{\partial p_i} = C(p_i)_m \operatorname{cosec} U_{mk}'. \quad \dots\dots (14)$$

Corresponding expressions follow for the transfer coefficients for the zonal and paraxial intersection points Z and P_x.

The intersection point Ab. For this point there are changes in two coordinate

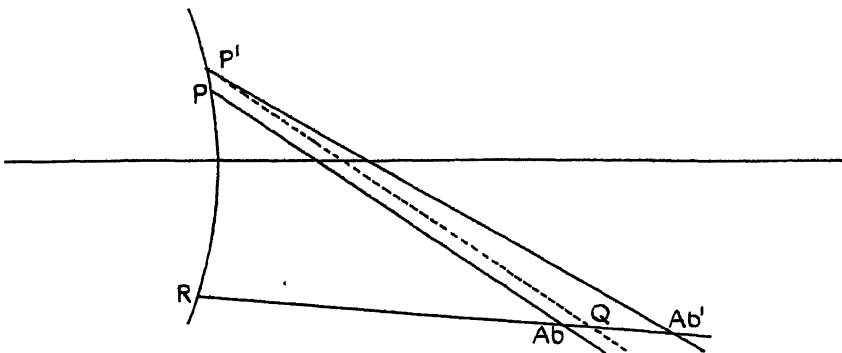


Figure 4.

to be considered. We begin by considering the effect of the change, dU_m' , made in the path of the ray a at surface i . In figure 4, P Ab and R Ab represent the

emergent traced paths of the rays a and b before any change is made. As a result of the change, dU'_{ai} , the ray a will now leave the last surface of the system along a new path $P'Ab'$, the point Ab' being the new position of the intersection point. Relative to the emergent traced path, the new ray has an incidence-point displacement, $dp'_k = (\partial p'_k / \partial U'_i)_a \cdot dU'_{ai}$, and a direction change, $dU'_k = (\partial U'_k / \partial U'_i)_a \cdot dU'_{ai}$. For small changes we have, to first order,

$$Ab Ab' = Ab Q + Q Ab' = (dp'_k - S'_{ak} dU'_k) \operatorname{cosec} (U'_a - U'_b)_k.$$

Resolving this displacement along directions parallel to and perpendicular to the principal axis, and differentiating with respect to U'_{ai} , we obtain

$$\begin{aligned} \frac{\partial L'_{ab}}{\partial U'_{ai}} &= \left(\frac{\partial p'_k}{\partial U'_{ai}} - S'_{ak} \frac{\partial U'_k}{\partial U'_{ai}} \right) \operatorname{cosec} (U'_a - U'_b)_k \cos U'_{bk} \\ &= C(U'_i)_a \operatorname{cosec} (U'_a - U'_b)_k \cos U'_{bk} \end{aligned} \quad \dots\dots(15)$$

$$\text{and} \quad \frac{\partial H'_{ab}}{\partial U'_{ai}} = -C(U'_i)_a \operatorname{cosec} (U'_a - U'_b)_k \sin U'_{bk}. \quad \dots\dots(16)$$

Similarly, a change, dp_{ai} , made in the path of ray a at surface i will lead to the expressions

$$\frac{\partial L'_{ab}}{\partial p_{ai}} = C(p_i)_a \operatorname{cosec} (U'_a - U'_b)_k \cos U'_{bk}, \quad \dots\dots(17)$$

$$\frac{\partial H'_{ab}}{\partial p_{ai}} = -C(p_i)_a \operatorname{cosec} (U'_a - U'_b)_k \sin U'_{bk}. \quad \dots\dots(18)$$

Turning now to the effects of changes dU'_{bi} , dp_{bi} , made in the path of the ray b at surface i , we can write down the final expressions for the transfer coefficients by inspection, a glance at figure 4 being sufficient to confirm the change of sign which occurs. Thus we have,

$$\frac{\partial L'_{ab}}{\partial U'_{bi}} = -C(U'_i)_b \operatorname{cosec} (U'_a - U'_b)_k \cos U'_{ak}, \quad \dots\dots(19)$$

$$\frac{\partial H'_{ab}}{\partial U'_{bi}} = C(U'_i)_b \operatorname{cosec} (U'_a - U'_b)_k \sin U'_{ak}, \quad \dots\dots(20)$$

$$\frac{\partial L'_{ab}}{\partial p_{bi}} = -C(p_i)_b \operatorname{cosec} (U'_a - U'_b)_k \cos U'_{ak}, \quad \dots\dots(21)$$

$$\frac{\partial H'_{ab}}{\partial p_{bi}} = C(p_i)_b \operatorname{cosec} (U'_a - U'_b)_k \sin U'_{ak}. \quad \dots\dots(22)$$

The intersection points Pr , Pr' . In figure 5, the traced path of the principal ray after leaving the last surface of the system is represented by RPr , and the corresponding path after a change dU'_i has been made at surface i is represented by $R'Pr'$. Relative to the traced path, the new emergent ray has an incidence-point displacement, $dp'_k = (\partial p'_k / \partial U'_i) dU'_i$, and a direction change, $dU'_k = (\partial U'_k / \partial U'_i) dU'_i$. Following the general lines of the previous cases, we have

$$Pr Pr' = Pr Q + Q Pr' = (dp'_k - S'_{kr} dU'_k) \sec U'_{prk},$$

from which it follows that

$$\begin{aligned} \frac{\partial H'_{pr}}{\partial U'_i} &= \left(\frac{\partial p'_k}{\partial U'_i} - S'_{kr} \frac{\partial U'_k}{\partial U'_i} \right) \sec U'_{prk} \\ &= C(U'_i)_{pr} \sec U'_{prk}. \end{aligned} \quad \dots\dots(23)$$

The corresponding expression for small dp -changes will be

$$\frac{\partial H'_{pr}}{\partial p_i} = C(p_i)_{pr} \sec U'_{prk}. \quad (24)$$

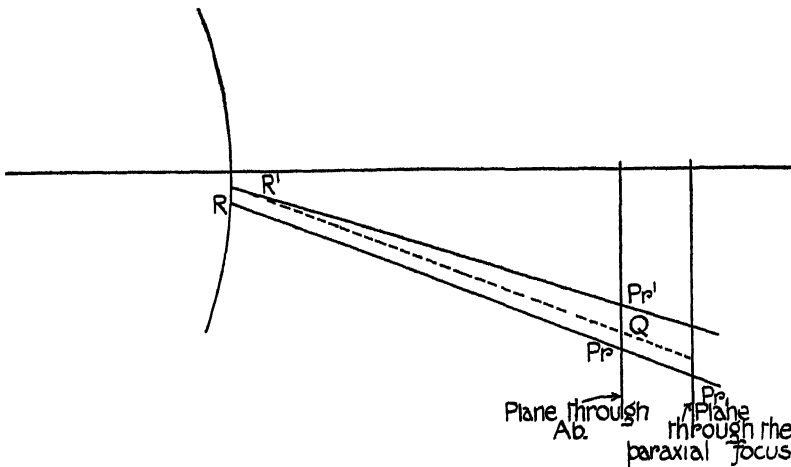


Figure 5

The transfer coefficients for the point Prf with respect to the changes dU'_i , dp_i , made in the principal ray at surface i will obviously be,

$$\frac{\partial H'_{prf}}{\partial U'_i} = C(U'_i)_{prf} \sec U'_{prk}, \quad \dots\dots (25)$$

$$\frac{\partial H'_{prf}}{\partial p_i} = C(p_i)_{prf} \sec U'_{prk}. \quad \dots\dots (26)$$

§ 5. THE INTERSECTION-POINT TRANSFER COEFFICIENTS FOR CHANGES OF CURVATURE, REFRACTIVE INDEX AND AXIAL SEPARATION

In the actual process of completing a design, the variables at the disposal of the designer are curvature, refractive index and the axial separation of successive surfaces. We now develop transfer coefficients for changes in these quantities.

Change of surface curvature. Suppose that a small curvature change, dc_i , is made at surface i of the system. To first order this leaves unchanged the incidence points of rays which strike the surface, and the total effect for any ray whose path has been traced through the original system is that after refraction the path of the ray will now be inclined to the traced path at an angle $dU'_i = (\partial U'_i / \partial c_i) dc_i$. Using the relations developed in the previous section, we can write down the new transfer coefficients at once.

For an axial ray, taking the marginal ray as typical, we have

$$\begin{aligned} \frac{\partial L'_m}{\partial c_i} &= \frac{\partial L'_m}{\partial U'_i} \cdot \frac{\partial U'_i}{\partial c_i} = \frac{\partial U'_i}{\partial c_i} C(U'_i)_m \operatorname{cosec} U'_{mk} \\ &= C(c_i)_m \operatorname{cosec} U'_{mk}. \end{aligned} \quad (27)$$

For the intersection point Ab it follows that since the a and b rays are affected by different amounts,

$$\begin{aligned}\frac{\partial L'_{ab}}{\partial c_i} dc_i &= dL'_{ab} = (dL'_{ab})_a + (dL'_{ab})_b \\ &= \frac{\partial L'_{ab}}{\partial U'_a} \frac{\partial U'_a}{\partial c_i} dc_i + \frac{\partial L'_{ab}}{\partial U'_b} \frac{\partial U'_b}{\partial c_i} dc_i,\end{aligned}$$

which on combination with (15) and (19) gives

$$\frac{\partial L'_{ab}}{\partial c_i} = [C(c_i)_a \cos U'_{bk} - C(c_i)_b \cos U'_{ak}] \operatorname{cosec} (U'_a - U'_b)_k, \dots\dots (28)$$

while the corresponding expression for $\partial H'_{ab}/\partial c_i$ is

$$\frac{\partial H'_{ab}}{\partial c_i} = -[C(c_i)_a \sin U'_{bk} - C(c_i)_b \sin U'_{ak}] \operatorname{cosec} (U'_a - U'_b)_k, \dots\dots (29)$$

We note that the positions of the intersection points Pr, Prf will be affected both by the change in the path of the principal ray and by the shift along the axis of the plane through Ab or the plane through Px. Thus, for a small change dc_i , we have

$$dH'_{pr} = \frac{\partial H'_{pr}}{\partial U'_i} \frac{\partial U'_i}{\partial c_i} dc_i - \frac{\partial L'_{ab}}{\partial c_i} dc_i \tan U'_{prk},$$

whence

$$\begin{aligned}\frac{\partial H'_{pr}}{\partial c_i} &= \frac{\partial H'_{pr}}{\partial U'_i} \frac{\partial U'_i}{\partial c_i} - \frac{\partial L'_{ab}}{\partial c_i} \tan U'_{prk} \\ &= C(c_i)_{pr} \sec U'_{prk} - \frac{\partial L'_{ab}}{\partial c_i} \tan U'_{prk}, \dots\dots (30)\end{aligned}$$

while the corresponding expression for the point Prf is

$$\frac{\partial H'_{prf}}{\partial c_i} = C(c_i)_{prf} \sec U'_{prk} - \frac{\partial l'}{\partial c_i} \tan U'_{prk}, \dots\dots (31)$$

Throughout these equations we have used the quantity $C(c_i)$ with the meaning defined by

$$C(c_i) = C(U'_i) \frac{\partial U'_i}{\partial c_i}. \dots\dots (32)$$

There remains only the ideal image point to consider. For an infinitely distant object point, for example, we have

$$H'_{id} = -f' \tan U_{pr} = -(y/u'_k) \tan U_{pr},$$

whence

$$\frac{\partial H'_{id}}{\partial c_i} = (y/u'_k{}^2) \tan U_{pr} \frac{\partial u'_k}{\partial c_i} \frac{\partial u'_i}{\partial c_i}. \dots\dots (33)$$

The differential coefficients in (33) are those normally calculated for the paraxial ray, so that the point Id fits conveniently into the general scheme.

Changes in refractive index. If the relative refractive index, $n=N/N'$, is changed by an amount dn_i at surface i , the ray after refraction at this surface will undergo a direction change $dU'_i = (\partial U'_i/\partial n_i) dn_i$, relative to the traced ray path. The transfer coefficients for the intersection points with respect to n -changes may therefore be written down from the corresponding expressions for the curvature

coefficients by replacing c everywhere by n . The C -quantity involved, $C(n_i)$, is given by

$$C(n_i) = C(U_i) \frac{\partial U_i'}{\partial n_i}. \quad \dots (34)$$

Changes in axial separation. The axial separation between the surfaces $(i-1)$ and i is specified by d_i . This quantity is essentially negative according to our sign convention, but it is more convenient than d_{i-1}' on account of the form of the final equations. In order to change the separation by an amount δd_i , we move the surface i and all succeeding surfaces through δd_i . Any ray incident on the surface i will meet the surface at a new incidence point after the change. Thus if the dotted arc in figure 6 represents the new position of the surface, the

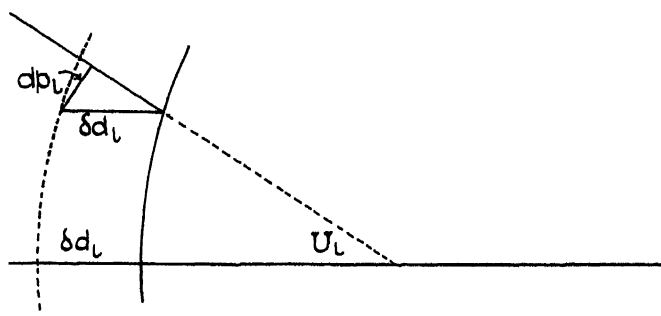


Figure 6

traced ray now meets the surface at a new incidence point given relative to the original by $\delta p_i = \sin U_i \delta d_i$. Hence the effect of a change in axial separation may be described in terms of incidence-point changes only. It follows that the expressions for the transfer coefficients for the intersection points with respect to d -changes can be obtained from those for the curvature changes by writing d everywhere in place of c . The C -quantity which occurs in these expressions is

$$C(d_i) = C(p_i) \frac{\partial p_i}{\partial d_i} = C(p_i) \sin U_i. \quad (35)$$

§ 6 GLASS CHANGE FOR A COMPONENT OF A LENS SYSTEM

In treating the refractive index as one of the variables of the system we have derived transfer coefficients which measure the aberration changes per unit change of the relative refractive index at each surface. In practice, a change of glass affects the complete component and thus involves two surfaces. It is necessary to develop a transfer coefficient which takes account of the complete change.

Let us denote the successive components of the system by subscripts a, b, \dots, h , using subscript h for the general component. It will be sufficient to denote the first and second surfaces of this general component by subscripts 1 and 2. The refractive index of the component is N_h , while the refractive indices of the media preceding and following the component will be denoted by N_{h-1} and N_{h+1} . If the glass of the component is now changed, its new index being

$N_h + dN_h$, the change in the value of n at the first surface is

$$dn_1 = \frac{\partial}{\partial N_h} \left(\frac{N_{h-1}}{N_h} \right) dN_h = - \frac{N_{h-1}}{N_h^2} dN_h, \quad \dots \dots (36)$$

while the change at the second surface is

$$dn_2 = \frac{\partial}{\partial N_h} \left(\frac{N_h}{N_{h+1}} \right) dN_h = dN_h / N_{h+1}. \quad \dots \dots (37)$$

The monochromatic aberrations. Excluding the chromatic aberrations Lch' and Tch' , which are treated separately, we can now write down the change in any aberration due to the glass change dN_h . Taking spherical aberration as typical, we have

$$dLA' = \frac{\partial LA'}{\partial N_h} dN_h = \frac{\partial LA'}{\partial n_1} dn_1 + \frac{\partial LA'}{\partial n_2} dn_2,$$

which, on combination with (36) and (37), gives

$$\frac{\partial LA'}{\partial N_h} = \frac{\partial LA'}{\partial n_2} \frac{1}{N_{h+1}} - \frac{\partial LA'}{\partial n_1} \frac{N_{h-1}}{N_h^2}, \quad \dots \dots (38)$$

with corresponding expressions for the other aberrations.

The chromatic aberrations. It is usual to adjust the longitudinal chromatic aberration to zero for some zone of the system, often the 0 707 zone. Suppose that the system is to be achromatized for two colours r and v . We write down expressions for the final intersection lengths of rays of these colours, incident at the selected zone, in terms of the intersection length of a traced ray of an intermediate colour d . Using first-order terms only, we have,

$$L_r' = L_d' + \frac{\partial L_d'}{\partial N_a} (N_r - N_d)_a + \frac{\partial L_d'}{\partial N_b} (N_r - N_d)_b + \dots \dots,$$

the sum extending over all the components of the system. Similarly,

$$L_v' = L_d' + \frac{\partial L_d'}{\partial N_a} (N_v - N_d)_a + \frac{\partial L_d'}{\partial N_b} (N_v - N_d)_b + \dots \dots$$

Hence
$$L_r' - L_v' = - \frac{\partial L_d'}{\partial N_a} (N_v - N_r)_a - \frac{\partial L_d'}{\partial N_b} (N_v - N_r)_b - \dots \dots,$$

that is,
$$Lch' = - \sum \frac{\partial L_d'}{\partial N_h} P_h, \quad \dots \dots (39)$$

where P_h is the partial dispersion $(N_v - N_r)_h$. If we change some of the glasses of the system so that the glass constants are altered by small amounts, we have, to first order,

$$dLch' = - \sum \frac{\partial L_d'}{\partial N_h} dP_h, \quad \dots \dots (40)$$

which means that the appropriate transfer coefficients for Lch' are of the type

$$\frac{\partial Lch'}{\partial P_h} = - \frac{\partial L_d'}{\partial N_h}. \quad \dots \dots (41)$$

In a similar manner we can deduce a corresponding expression for the transverse chromatic aberration and appropriate transfer coefficients. A measure of the transverse chromatic aberration is provided by $H'_{pr} - H'_{prv}$, the intersection

points concerned being those in which the principal rays for colours r and v cut a fixed plane perpendicular to the axis, which is either the paraxial image plane or the plane through Ab. Using the coefficients for the traced principal ray of intermediate colour d , we have

$$H'_{prv} = H'_{prd} + \left(\frac{\partial H'_{prd}}{\partial N_a} \right)_{L'} (N_v - N_d)_a + \left(\frac{\partial H'_{prd}}{\partial N_b} \right)_{L'} (N_v - N_d)_b + \dots,$$

with a corresponding expression for H'_{prv} . The subscript L' indicates that the plane in which the intersection points occur remains fixed. Hence we have

$$H'_{prv} - H'_{prv} = - \left(\frac{\partial H'_{prd}}{\partial N_a} \right)_{L'} (N_v - N_r)_a - \left(\frac{\partial H'_{prd}}{\partial N_b} \right)_{L'} (N_v - N_r)_b - \dots,$$

that is,
$$Tch' = - \Sigma \left(\frac{\partial H'_{prd}}{\partial N_h} \right)_{L'} P_h, \quad \dots (42)$$

whence, also,
$$\frac{\partial Tch'}{\partial P_h} = - \left(\frac{\partial H'_{prd}}{\partial N_h} \right)_{L'}. \quad \dots (43)$$

The expression for $\left(\frac{\partial H'_{prd}}{\partial N_h} \right)_{L'}$ follows fairly obviously from (30), (34), and (38).

Equations (39) and (42) are of great value in practical designing. They provide a very rapid guide to glass changes which will improve the achromatism of the system, and permit an analysis of the distribution of the intersection lengths of rays of different colours without the labour of separate tracings. The order of accuracy is fairly high. As an example, the chromatic aberrations of a photographic lens of six components was calculated by (39) and (42), and checked by tracing, giving the following values:

	Lch'_s	lch	Tch'
From eqns. (39), (42)	0.429	0.393	0.00516
From trace	0.440	0.387	0.0059

the differences being insignificant when the uncertainty of the last figure in the five-figure trace is considered.

§7 THE TRANSFER COEFFICIENTS FOR SMALL CHANGES OF THE DIAPHRAGM POSITION

If the procedure developed for the alteration of the axial separation of successive surfaces is applied at the first surface of the system, there is no alteration in the spacings of the surfaces of the system. It remains, then, to interpret the meaning of the values of the transfer coefficients for d -changes which are obtained by the formal application of our equations at the first surface. Thinking of the incident traced rays in the space to the left of the first surface as fixed, a shift of the first surface will result in incidence-point changes for all rays except those incident parallel to the principal axis. In particular, the original principal ray of an oblique pencil will strike the first surface at a new incidence point, and will no longer pass through the centre of the diaphragm, but will intersect the principal axis at some neighbouring point. If the diaphragm is shifted so that its centre coincides with this point, the ray in its new course will be a principal ray for this

new diaphragm position. The surface shift, made in the manner considered, is therefore equivalent to a change in the position of the diaphragm. The transfer coefficients associated with the surface shift must therefore give information as to the effect of the shift of the diaphragm on the aberrations.

Suppose that the diaphragm is placed between the surfaces i and $(i+1)$ of the system, its axial distance from the pole of the first surface being L'_{pr} . The principal ray in a pencil of the original trace is aimed at the centre of the entrance pupil, O , distant L_{pr} from the pole of the first surface, and passes through O' , the centre of the diaphragm. If the point O is shifted a distance dL_{pr} , the point O' will be displaced through dL'_{pr} , and the ratio of these two displacements is the longitudinal magnification. As an approximation let us assume that the longitudinal magnification is equal to the square of the lateral magnification, as in paraxial theory, and we may write

$$\frac{dL'_{pr}}{dL_{pr}} = M'^2 = \frac{\sin^2 U_{pr}}{\sin^2 U'_{p,i}}. \quad \dots\dots(44)$$

A shift of the first surface, carried out in the manner prescribed, is equivalent to an equal shift of the entrance pupil. Hence, for any aberration A ,

$$\frac{\partial A}{\partial L_{pr}} = \frac{\partial A}{\partial d_1} \frac{\partial L_{pr}}{\partial L'_{pr}} = \frac{\partial A}{\partial d_1} \frac{\sin^2 U'_{p,i}}{\sin^2 U_{pr}}, \quad \dots\dots(45)$$

giving a series of transfer coefficients for small displacements of the diaphragm.

Another useful feature incidental to the transfer coefficients of the principal ray at the first surface of the system is that the t -focus is readily available from them, thus eliminating the necessity of a trace for close tangential rays. If a small dp -change is made for the principal ray at the first surface, the resulting dp'_k and dU'_k specify the emergence point and direction of a ray close to the principal ray as it leaves the last surface of the system. It is immediately obvious that the intersection of this ray with the principal ray gives the t -focus, that is,

$$t'_k = \frac{\partial p'_k / \partial p_1}{\partial U'_k / \partial p_1}. \quad \dots\dots(45 a)$$

REFERENCES

- CONRADY, 1929 *Applied Optics and Optical Design* (Oxford . Clarendon Press).
M'AULAY and CRUICKSHANK, 1945. *Proc. Phys Soc* 57, 302.

A SYSTEM OF TRANSFER COEFFICIENTS FOR USE IN THE DESIGN OF LENS SYSTEMS: II. A SECOND-ORDER CORRECTION TERM

By F. D. CRUICKSHANK,

University of Tasmania

MIS received 25 January 1944; in revised form 26 February 1945

ABSTRACT Expressions are derived for a second-order correction term for the differential transfer coefficients developed in Part I. Need for the correction term arises on account of the neglect of the effect of the small change in the direction of the emergent ray during the calculation of the transfer coefficients for the intersection points in the final image-space. The correction term is used when the effects of changes in the system beyond differential size are required with increased accuracy. For single changes of this kind in a system, the correction term frequently improves the accuracy from the order of 7 to 10 per cent to less than one per cent. The necessary modifications are given for the case in which a series of changes are made simultaneously within the system.

§ 1 INTRODUCTION

IN the account of the transfer coefficients given in Part I we have considered only terms of first order, a procedure which is adequate for general designing needs as long as small changes only are made in the system. In the course of correcting a system it is frequently necessary to make curvature changes of the order of 0.001 mm.^{-1} , and thickness changes of the order of one millimetre. When such changes are made, especially in systems of large aperture, it is found that the shifts of the various intersection points in the final image space as predicted by the transfer coefficients may be in error by as much as twelve per cent. There is one second-order term associated with the final stage of the calculation of the transfer coefficients, the use of which will improve considerably the order of accuracy of such changes. It arises from the neglect of the effect of the small change, dU_k' , in the final direction of the emergent ray, on the computation of the transfer coefficients for the intersection points from the corresponding C -quantities. It would seem that the wider problem of second-order terms affecting the fundamental transfer coefficients is such that the amount of computation involved is incommensurate with the results obtained therefrom.

§ 2. THE CORRECTION TERM FOR SINGLE CHANGES MADE WITHIN THE SYSTEM

In figure 7 we represent by R_1M a traced ray in the final image space which intersects the optical axis of the system at M . As a result of a change made within the system, let us say a curvature change, dc_i , at surface i , the emergent ray now follows the path R_2M' . The lines MP and MQ are drawn perpendicular to R_1M and R_2M' respectively. Then

$$\begin{aligned} MP &= dp_k' - S_k' dU_k' = C(c_i) dc_i, \\ MQ &= MP \cos dU_k', \\ MM' &= MQ \operatorname{cosec} (U_k' + dU_k') \\ &= C(c_i) dc_i \cos dU_k' \left[\operatorname{cosec} U_k' + \frac{\partial}{\partial U_k'} (\operatorname{cosec} U_k') dU_k' \right]. \end{aligned}$$

It would be most unusual for dU_k' to be greater than about 2° , so that we may safely put its cosine = 1. Writing $MM' = \delta L_k'$, we have

$$\delta L_k' = C(c_i)dc_i \operatorname{cosec} U_k'(1 - \cot U_k' dU_k') \quad \dots\dots(46)$$

$$= C(c_i) \operatorname{cosec} U_k' dc_i - C(c_i) \operatorname{cosec} U_k' \cot U_k' \frac{\partial U_k'}{\partial c_i} (dc_i)^2. \quad \dots\dots(47)$$

The first term on the right-hand side of (47) is the first-order term which we have

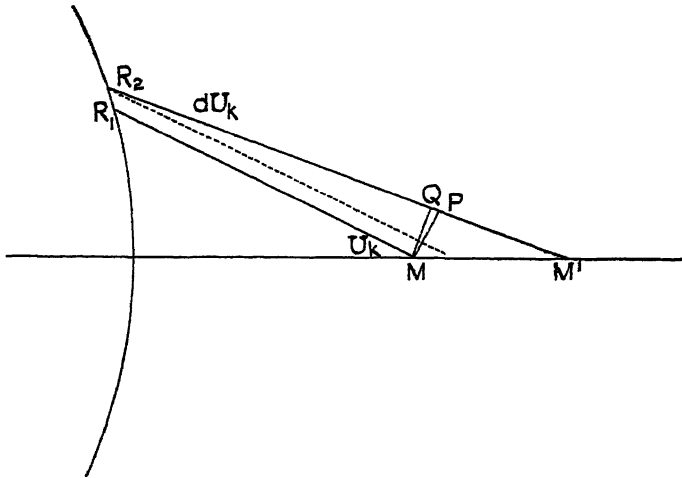


Figure 7.

used all along, while the second term now appears as a correction to be added to it. We write this correction term as $\lambda'(c_i)(dc_i)^2$, where

$$\lambda'(c_i) = -C(c_i) \operatorname{cosec} U_k' \cot U_k' \frac{\partial U_k'}{\partial c_i}. \quad \dots\dots(48)$$

In the next place we consider the intersection of a ray in the final image space with a fixed plane at right angles to the optical axis. Proceeding in a manner similar to the foregoing, it is easily seen that the change in the H' -coordinate of the intersection point is given by

$$\begin{aligned} \delta H_k' &= C(c_i)dc_i \cos dU_k' \sec(U_k' + dU_k') \\ &= C(c_i)dc_i \sec U_k'(1 + \tan U_k' dU_k') \quad \dots\dots(49) \end{aligned}$$

$$= C(c_i) \sec U_k' dc_i + C(c_i) \sec U_k' \tan U_k' \frac{\partial U_k'}{\partial c_i} (dc_i)^2. \quad \dots\dots(50)$$

The first term on the right-hand side of (50) is the usual first-order term, while the second is the correction term to be applied where greater accuracy is desired. Writing this in the form $\nu'(c_i)(dc_i)^2$, we have

$$\nu'(c_i) = C(c_i) \sec U_k' \tan U_k' \frac{\partial U_k'}{\partial c_i}. \quad \dots\dots(51)$$

There remains one further important case to consider, that in which the ray intersects a plane at right angles to the axis of the system, the plane being no longer fixed, but changing its position when an alteration is made within the system. Having in mind the use to be made in subsequent work of intersections

of rays with the paraxial image plane, we will consider this case in detail. By analogy with equation (31), Part I, it follows that for any ray intersecting the paraxial image-plane

$$\frac{\partial H_k'}{\partial c_i} = C(c_i) \sec U_k' - \frac{\partial l_k'}{\partial c_i} \tan U_k'.$$

Extending this expression to take account of the effect of the change in U_k' , it is fairly obvious that for the change in the H' -coordinate of the intersection point resulting from the curvature change, dc_i , we shall have

$$\begin{aligned} \delta H_k' &= C(c_i) dc_i \sec (U_k' + dU_k') \\ &\quad - \left[\frac{\partial l_k'}{\partial c_i} dc_i + \lambda_0'(c_i) (dc_i)^2 \right] \tan (U_k' + dU_k'), \quad \dots\dots (52) \end{aligned}$$

where $\lambda_0'(c_i)$ is the λ' correction term for the paraxial ray, which, by analogy with (48), is seen to be

$$\lambda_0'(c_i) = - \frac{\partial l_k'}{\partial c_i} \frac{1}{u_k'} \frac{\partial u_k'}{\partial c_i}. \quad \dots\dots (53)$$

Developing the terms of equation (52), we obtain

$$\begin{aligned} \delta H_k' &= C(c_i) dc_i \sec U_k' (1 + \tan U_k' dU_k') \\ &\quad - \left[\frac{\partial l_k'}{\partial c_i} dc_i + \lambda_0'(c_i) (dc_i)^2 \right] \tan U_k' (1 + \operatorname{cosec} U_k' \sec U_k' dU_k') \\ &\quad \dots\dots (54) \end{aligned}$$

Multiplying out the bracketed expressions and neglecting the third-order term which occurs, we obtain

$$\delta H_k' = \text{first-order terms} + [\nu'(c_i) + \gamma'(c_i) + \rho'(c_i)] (dc_i)^2$$

where we have written

$$\nu'(c_i) = - \frac{\partial l_k'}{\partial c_i} \tan U_k' \operatorname{cosec} U_k' \sec U_k' \frac{\partial U_k'}{\partial c_i} \quad \dots\dots (55)$$

and

$$\begin{aligned} \rho'(c_i) &= -\lambda_0'(c_i) \tan U_k' \\ &= \frac{\partial l_k'}{\partial c_i} \tan U_k' \frac{1}{u_k'} \frac{\partial u_k'}{\partial c_i}. \quad \dots\dots (56) \end{aligned}$$

We may now write the correction term which is to be applied for intersection points in the paraxial image-plane as $\mu'(c_i) (dc_i)^2$, where

$$\mu'(c_i) = \nu'(c_i) + \gamma'(c_i) + \rho'(c_i). \quad \dots\dots (57)$$

In spite of the apparent complexity of these expressions, they are computed very easily on account of the fact that portions of each term have already been evaluated at an earlier stage in the general computation. Some examples taken at random from a computation are set out below, the system being an f/1.5 camera lens of focal length 106.5 mm. comprising ten surfaces. These examples will afford some idea of the effect of the use of the correction terms for single changes made within the system.

Example 1. The extreme marginal ray of the axial pencil of this lens had the following coefficients for curvature changes at the first surface:

$$\frac{\partial L_k'}{\partial c_1} = -9913.0, \quad \lambda'(c_1) = 878530.$$

On making a change of curvature at the first surface of amount 0.001 mm.^{-1} ,

the radius thereby changing from +72.84 mm. to +67.89 mm., the following results are obtained :

Expected change in L_k' using first-order coefficient = -9.913 mm.

Expected change corrected by second-order term = -9.034 mm.

Change as calculated by trace = -8.838 mm.

The use of the correction term in this case reduces the error of the predicted change from 12.16 per cent to 2.22 per cent.

Example 2. A paraxial ray traced through the same lens had the following coefficients at the first surface :

$$\frac{\partial L_k'}{\partial c_1} = -7447.2, \quad \lambda_0'(c_1) = 446350.$$

For the same change in curvature at the first surface we find :

Expected change in L_k' by first-order coefficient = -7.447 mm.

Expected change, corrected = -7.001 mm.

Change as given by trace = -7.031 mm.

In this case the use of the correction term reduces the error of the predicted change from 6.35 per cent to 0.43 per cent.

Example 3. The principal ray of an oblique pencil incident on the same system at an angle of 17° to the principal axis had the following coefficients :

$$\frac{\partial H_k'}{\partial c_1} = 1764.8, \quad \mu'(c_1) = -117550.$$

For the same curvature change made at the first surface of the system we have :

Expected change in H_k' by first-order coefficient = 1.765 mm.

Expected change, corrected = 1.647 mm.

Change as calculated by trace = 1.637 mm.

The error in the prediction is thus reduced from 7.82 to 0.61 per cent.

§ 3. THE CORRECTION TERM FOR SEVERAL SIMULTANEOUS CHANGES

In the course of an actual design it is usual to make several changes in the system at the same time and then to check the result either in full or in part by a new trace of the system. If the changes are such that a considerable alteration has been made in the positions of the intersection points, it will generally be desirable to use the dU_k' correction term to improve the accuracy of the expected changes. In order to do this we must alter the form of the expressions for these correction terms. In applying the first-order theory to several changes made simultaneously, we treat the change in position of an intersection point as a total differential. Thus, suppose that changes are made in the curvatures of surfaces i, j, \dots, r , then for an intersection point on the principal axis

$$\begin{aligned} dL_k' &= \frac{\partial L_k'}{\partial c_i} dc_i + \frac{\partial L_k'}{\partial c_j} dc_j + \dots + \frac{\partial L_k'}{\partial c_r} dc_r \\ &= \sum_i \left(\frac{\partial p_k'}{\partial c_i} - S_k' \frac{\partial U_k'}{\partial c_i} \right) \operatorname{cosec} U_k' dc_i \\ &= (\sum dp_k' - S_k' \sum dU_k') \operatorname{cosec} U_k'. \end{aligned} \quad \dots (58)$$

The summation of the effects of the several differential changes is to be made, therefore, at the last surface of the system, so that the ray under consideration will emerge from the last surface with an incidence-point displacement, $\Sigma dp_k'$, and a direction change, $\Sigma dU_k'$, relative to the original traced path. For changes that are not differential in size we still obtain a useful tool for design by assuming that the summation property holds, and the order of accuracy can be improved by applying the correction term. The value of the small angle-change, dU_k' , for which the correction is made, is of course the sum of the various dU_k' due to the several changes. The correction terms must therefore be written as factors which multiply the resultant angle change, $\Sigma dU_k'$, and not as factors multiplying $(dc_i)^2$, etc., as in the previous section. Anticipating the needs of subsequent methods, we consider the two main cases, the intersection point of the paraxial ray with the principal axis and the intersection point of any ray with the paraxial image-plane.

In the former case it is easily seen that by analogy with equation (46)

$$\begin{aligned}\delta l_k' &= \Sigma \frac{\partial l_k'}{\partial c_i} dc_i \left(1 - \frac{1}{u_k'} \Sigma du_k'\right) \\ &= \Sigma \frac{\partial l_k'}{\partial c_i} dc_i - \left(\Sigma \frac{\partial l_k'}{\partial c_i} \frac{1}{u_k'} dc_i\right) (\Sigma du_k').\end{aligned}\quad \dots\dots(59)$$

Hence the correction term to be applied for the paraxial intersection point is $(\Sigma \lambda_0(c_i) dc_i)(\Sigma du_k')$, where

$$\lambda_0(c_i) = - \frac{\partial l_k'}{\partial c} \frac{1}{u_k'} \quad \dots\dots(60)$$

and

$$\Sigma du_k' = \Sigma \frac{\partial u_k'}{\partial c_i} dc_i \quad \dots\dots(61)$$

In the case of the intersection of any ray with the paraxial image-plane we have by analogy with equation (52)

$$\begin{aligned}\delta H_k' &= [\Sigma C(c_i) dc_i] \sec(U_k' + \Sigma dU_k') \\ &\quad - \left[\Sigma \frac{\partial l_k'}{\partial c_i} dc_i + \Sigma \lambda_0(c_i) dc_i \Sigma du_k' \right] \tan(U_k' + \Sigma dU_k') \\ &= \Sigma \left[C(c_i) \sec U_k' - \frac{\partial l_k'}{\partial c_i} \tan U_k' \right] dc_i \\ &\quad + \Sigma C(c_i) \sec U_k' dc_i \tan U_k' \cdot \Sigma dU_k' \\ &\quad - \Sigma \frac{\partial l_k'}{\partial c_i} \tan U_k' dc_i \operatorname{cosec} U_k' \sec U_k' \cdot \Sigma dU_k' \\ &\quad + \Sigma \frac{\partial l_k'}{\partial c_i} \tan U_k' dc_i \frac{1}{u_k'} \cdot \Sigma du_k' \\ &= \text{first-order term} + \Sigma \nu(c_i) dc_i \cdot \Sigma dU_k' \\ &\quad + \Sigma \gamma(c_i) dc_i \cdot \Sigma dU_k' + \Sigma \rho(c_i) dc_i \cdot \Sigma du_k',\end{aligned}\quad \dots\dots(62)$$

where

$$\nu(c_i) = C(c_i) \sec U_k' \tan U_k', \quad \dots\dots(63)$$

$$\gamma(c_i) = - \frac{\partial l_k'}{\partial c_i} \tan U_k' \sec U_k' \operatorname{cosec} U_k', \quad \dots\dots(64)$$

$$\rho(c_i) = \frac{\partial l'_k}{\partial c_i} \tan U'_k \frac{1}{u'_k}. \quad \dots\dots(65)$$

Combining the first two of these terms by writing

$$\pi(c_i) = \nu(c_i) + \gamma(c_i), \quad \dots\dots(66)$$

the correction term becomes

$$\Sigma \pi(c_i) dc_i \cdot \Sigma dU'_k + \Sigma \rho(c_i) dc_i \cdot \Sigma du'_k.$$

It will have been noted that the quantities corresponding to $\nu(c_i)$, $\gamma(c_i)$, $\rho(c_i)$ and $\lambda(c_i)$ used in the first section are distinguished from those of the present section by the addition of a dash. This serves to remind the computer that they relate to single changes only. It is to be stressed that the correction term for a series of changes is only of use when there is a considerable change of the intersection points in the image space. During the correction of a design it is not unusual to make up to eight or ten changes, the total result of which may be to displace the intersection points in general through quite small distances, although the small relative shifts of these points may alter the aberrations considerably. However, when the result of the series of changes shifts the intersection points through distances of the order of 10% of the focal length of the system, the effect of the correction term is worth while

We have not considered the correction terms for changes of thickness or refractive index, as these may be written down by analogy with the curvature expressions.

REVIEWS OF BOOKS

The Measurement of Colour, by W. D. WRIGHT. Pp. vii+223. (London: Adam Hilger, Ltd., 1944.) 30s.

The main theme of this book is the trichromatic system of specifying colours, and the exposition of the principal facts and formulae of the system occupies about one-fifth of the whole. The emission, reflection and transmission of radiation and the response of the eye to light are dealt with in introductory chapters. The rest of the book is devoted to the practical methods of colorimetry and spectrophotometry, both visual and photo-electric, and to applications of colour measurement in industry, agriculture, medicine, etc.

Dr Wright writes with authority in this field, and of the soundness of the treatment it need only be said that it is as expected. Great pains are taken to make the subject intelligible to those who are not familiar with linear transformation theory, projective and differential geometry, and the basic ideas are well illustrated by analogies and coloured diagrams. The result is an excellent elementary presentation of trichromatic colorimetry. Of the few points which may be criticized, the first is the author's discussion of the use of the visibility curve in fixing the distribution coefficients of the colorimetric observer. This was an expedient imposed by the fact that in neither of the fundamental determinations of the mixture data were measurements made of the energy in the spectrum. Wright's treatment here is a little confusing, as he links this unfortunate gap in our knowledge with the difficulty of designing a colorimeter incorporating the reference stimuli as instrumental stimuli. But had such a colorimeter been used and no energy measurements made, the visibility curve would still have been required, and conversely with the energy measurements made, the visibility curve could have been dispensed with, whatever the instrument stimuli. While variations of the colour matches of different observers are stressed in several places, no indication of their magnitude is given. A plot of the mixture curves through the spectrum for a group of observers and Wright's diagram showing how differences in retinal pigmentation scatter the "white points" in a chromaticity chart based on spectral primaries might have been included with advantage. In the chapter on colour atlases a brief explanation of the Oswald system, similar to the excellent summary of the Munsell system, would have been welcome.

Admirable features are the concise explanation of the principles underlying the many measuring instruments used in colour work, and the clear statement in simple language of colour problems in the dyeing and paint industries, in colour reproduction and in photo-elasticity. The book should be of particular value to those newly entering this field of work, while all concerned with colour questions will find it interesting reading. It is a pity the price is so high.

W. S. S.

Tables of Elementary Functions, by F. EMDE. Pp. xii+181 (Leipzig: Teubner, 1940; photographic reprint published under licence by J. W. Edwards, Ann Arbor, Mich., U.S.A., 1945.) \$3.20.

Most physicists will be familiar with "Jahnke und Emde"; many will remember that the third (1938) edition omitted the material of the first 76 pages of the second (1933) edition, and that an extended set of tables of the "elementary" functions was promised. The volume appeared in Germany in 1940, and this reprint makes it generally available.

We have the familiar mixture of letterpress and formulae, numerical tables and diagrams. The above-mentioned material has all been incorporated, but a larger page has been used, and all the numerical tables have been reset, using the "heads and tails" numerals, which make them so much easier to read. Four-figure accuracy is standard, and the tabular intervals are generally so chosen that it is attainable by linear interpolation (for which a slide-rule is advocated). Divided differences are given, and in some ranges where linear interpolation is not adequate they are printed in italics. Text and legends are in both German and English.

Space permits the mention of only the more important additions. We have x^2 and $1/x$, x^3 and $1/x^3$, and square and cube roots, in the section on powers. 100 multiples of

$M(=\log_{10} e)$, $1/M$, $\pi/2$ and $2/\pi$ are given to six decimals, and 10 multiples to 16 decimals. A factor table includes all composite odd numbers less than 10,000 not divisible by 3 or 5. There is a section, with numerical tables, on quadratic equations—in which a diagram similar to figure 11 for the cubic might with advantage have been included—and a long account of the quartic, with diagrams but no significant numerical material.

Probably the most important addition is the set of tables of circular, exponential, and hyperbolic functions. Sine, cosine, tangent and cotangent are given for angles in degrees. A comprehensive set of tables of the direct and inverse functions, in natural (radian) measure, is given. There is also a set of tables of the direct functions in terms of the quadrant as a unit, i.e., the circular, exponential and hyperbolic functions of the argument $\pi x/2$. There are also tables of a few functions such as $1 - \cos x$, $\cosh x - 1$, $\sqrt{2} \sin x$, $\sqrt{2} \cos x$.

Some special functions such as e^{-x^2} , $\coth x - 1/x$, $\tan(x\sqrt{i})/x\sqrt{i}$ (pertinent to skin effect), and some Chebyshev polynomials, are included. A section near the end gives the coefficients in powers and roots of polynomials (truncated series). There is also a very useful few pages of advice on computing, and a bibliography.

The book will be of great use to that body of engineers and physicists who have frequent dealings with the "elementary" functions, but only rarely need the "higher" ones. The compiler of such a book is obviously dependent, in the main, upon extant computations, so that scrappiness and unevenness are inevitable. Much of the apparent disorder of the arrangement is due to the idiosyncrasies of the individual functions tabulated. The important thing is that the information is in the book, and that with the help of a detailed list of contents and a comprehensive index it is not difficult to find. W. G. B.

The Velocity of Light. by N. ERNEST DORSEY. Pp. 110. (Trans. Amer. Phil. Soc. vol. xxiv, part 1, 1944.) \$2.25.

If a list of physical quantities were made in the order of the time and attention devoted to their measurement, most physicists would place the velocity of light very high in that list. It will come as something of a shock to learn that many of these "standard" investigations leave a great deal to be desired in view of the accuracy claimed.

The author examines, in detail, the investigations, using the toothed wheel (Fizeau, Cornu, Perrotin-Prim), the rotating mirror (Foucault, Newcomb, Michelson, Michelson-Pearse-Pearson) and the Kerr-cell methods (Karolus-Mittelstaedt, Anderson, Hüttel). Particular care is devoted to the consideration of possible systematic errors, to the evidence of a search for systematic errors on the part of the investigators, and to the standard of the reporting of all relevant data. The author's conclusions are summarized in the following table, taken from the paper.

Summary of values for the velocity of light in a vacuum
(unit of $V=1$ megameter/second in vacuum)

Observer and date	Range for V	Centre V	Remarks
Cornu 1872	296.5–300.5	298.5	Preliminary
Cornu 1876	299.3–300.5	299.9	Uncertain
Perrotin and Prim 1908	299–301	300	Systematic error
Newcomb 1880–82	299.71–299.86	299.78	Systematic error
Michelson 1878	297–304	300	Preliminary
Michelson 1879	299.7–300.1	299.9	Systematic error
Michelson 1882	299.6–300.1	299.85	Systematic error
Michelson 1924	299.73–299.87	299.80	Report deficient
Michelson 1927	299.77 ₈ –299.81 ₈	299.79 ₈	Report deficient
Michelson, Pearse, Pearson 1935	299.76 ₄ –299.78 ₄	299.77 ₄	Little study of apparatus
Karolus and Mittelstaedt 1929	299.75 ₈ –299.79 ₈	299.77 ₈	Apparatus studied
Anderson 1937	299.75 ₈ –299.78 ₈	299.77 ₁	Apparatus studied
Anderson 1941	299.76 ₈ –299.79 ₀	299.77 ₆	Apparatus studied
Hüttel 1940	299.75 ₈ –299.77 ₈	299.76 ₈	Apparatus studied

The last five of these determinations are alone considered to have other than historical interest. The best value for the velocity of light is given as 299,77₃ km./sec. in a vacuum, with a dubiety of ± 10 km/sec. The possibility of a secular variation in the velocity of light is dealt with briefly, with the conclusion that the data give no indication of any variation.

The report forms a valuable study in critical examination; it also contains a statement of the standard to which an accurate experimental investigation and its reporting ought to conform, which should be read once a year by every physicist for the good of his soul.

M. B.

School Physics, Part II, by T. M. YARWOOD. Pp. x+438. (London: Macmillan and Co., Ltd., 1945.) 7s. 6d. net.

To appraise the value of this *School Physics*, Part II, it is necessary to study it concurrently with Part I, since both books cover the same ground and form, as stated in the preface, "a four-year concentric school course in Physics whereby a certain amount of work in all branches of Physics is done in each year". Thus the Section headings, with the exception of an additional one on *Molecular Forces* in Part II, are identical for both volumes. The author is concerned to apply physical principles to everyday life, and he includes numerous references to modern appliances. Especially is he anxious to encourage the future citizen to become more air-minded. The book makes interesting reading with its good and simple diagrams and its generally clear explanations of working principles. It is carefully produced, though the reader of p. 259 might conclude that the overtones of most vibrating systems, including bells, were harmonic.

The subject-matter includes those portions of physics which may best be studied at school by the average pupil, so that this publication, at a time when school, scholarship and university syllabuses are in the melting pot, is very welcome. For effective teaching, the book would need to be supplemented by a historical and biographical background.

M. D. W.

Sechsstellige trigonometrische Tafeln, by H. BRANDENBURG. Pp. xxiv+304 (Leipzig: Lorentz, 1932; photographic reprint published under licence by J. W. Edwards, Ann Arbor, Mich., U.S.A., 1945.) \$5.00.

This is a reprint of a well-known table. It gives the sine, tangent, cotangent and cosine, for angles in a quadrant at interval 10 seconds, to six decimals (except for tangent and cotangent for a few degrees), with differences and proportional parts (pp. 30-299). Preliminary tables give the cotangent, for every second up to 3°, to at least seven figures (pp. 2-24), and sines and tangents, at interval 10 seconds up to 1°, to seven decimals (pp. 26-28). At the end are given the values of $n!$, $1/n!$, $(\pi/2)^n$ and $(\pi/2)^n/n!$ for $n=0(1)20$ (p. 300), tables for converting sidereal into mean solar time and vice versa (pp. 302-3), and for converting angles into time (p. 304).

The preface is given in seven languages.

Some errata (supplied by Dr. L. J. Comrie) are printed on p. iv

W G B.

CORRIGENDA

Page 231. The legend "Colours of sunlight and daylight" appearing under figure 5 should be transferred to p. 235, under figure 9.

Page 233. The point in figure 7 marked "SB(2) 20% Tm 24.1" should be marked "SB(2) 20° c. Tm 24.1".

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 57, PART 5

1 September 1945

No. 323

THE FRICTIONLESS STATE OF AGGREGATION

BY K. MENDELSSOHN,

Clarendon Laboratory, Oxford

MS. received 26 March 1945

ABSTRACT. The analogy of superconductivity and liquid helium II has been analysed on a purely empirical basis. Frictionless transport is considered to be the fundamental phenomenon. It has been concluded that the "superconductive" electrons and the "superfluid" helium atoms form an aggregate in momentum space of zero thermal energy (α -state). The α -particles occupy a set of lowest quantum states which is separated from the thermal states by an energy gap; they have zero entropy, but an appreciable zero-point energy. The observed specific-heat anomalies are due to a lifting of α -particles over this gap. The transport phenomena are caused by a diffusion of α -particles under their zero-point momentum. This zero-point diffusion does not result in dissipation of momentum. It explains why frictionless transport is independent of the strength of external forces and explains also the critical transport rates (threshold value and film transfer). Bose-Einstein condensation fails to account for these characteristic features because it does not lead to a zero-point energy of the condensed phase. A number of experiments to test the hypothesis are suggested.

§1 INTRODUCTION

SOME years ago J. G. Daunt and the author (1942) drew attention to the strong similarity between the phenomena of superconductivity and liquid helium II, which becomes apparent when we consider quite generally the properties of particle transport in these substances. In both cases we encounter *frictionless* transport which obeys similar rules in spite of the fact that the particles are atoms in one case and electrons in the other. This, together with an equally striking similarity in the entropy functions (Daunt and Mendelssohn, 1945) has led us to advance the hypothesis that both phenomena are caused by much the same mechanism. We have described this mechanism as the appearance, at finite temperature, of particles which are energetically at absolute zero and do not exchange momentum with the remainder of the substance. In view of the peculiar way in which the thermally unexcited particles are separated from those in the higher quantum states—they do not form a separate phase—we have described the assembly of these particles as an aggregation.

It is the aim of this paper to discuss in more general terms* the hypothesis of the basic identity of superconductivity and liquid helium, and the properties of the aggregation of particles at the lowest quantum levels. It must be emphasized that these considerations are not based on any particular theoretical conception

* A detailed discussion of the energy state of the electrons in superconductors has recently been given elsewhere (Daunt and Mendelssohn, 1945) and a detailed account of the corresponding phenomena in liquid helium will be published later.

but that an attempt is made to deduce the properties of the frictionless state empirically on the basis of the existing experimental evidence.

§ 2. FRICTIONLESS TRANSPORT

Let us first consider the position as regards the experimental facts.* In the case of liquid helium the last years before the war had brought to light a considerable number of seemingly unconnected phenomena, the most striking of which are : extremely low viscosity, very high heat conduction and, in some way related to the latter, a thermo-mechanical effect (sometimes called the "fountain effect") which constitutes a flow of helium in the direction of the higher temperature. In addition there are the transport phenomena in a film of thickness about 5×10^{-6} cm. which covers all solid surfaces in contact with liquid helium II. The somewhat unsatisfactory state of the experimental material is due to the fact that internal friction and heat transport in liquid helium II are evidently very complicated processes. The efforts to determine such properties as viscosity and heat conduction have consequently led to complex, and in many cases contradictory, results. On the other hand, the film transfer which at first sight appeared to complicate matters has proved to be of great help since the phenomena in films seem to present a much simplified picture of the transport effects in the bulk liquid and have the added advantage of yielding well reproducible results.

With regard to superconductivity, the position is more satisfactory on account of the greater wealth of confirmed experimental results. Since 1933, when it was discovered that zero resistance is coupled with zero magnetic induction, there has been a tendency to describe superconductivity as a purely magnetic phenomenon, some authors going so far as to consider the disappearance of resistance to be due to complete diamagnetic polarization. Such an explanation is, however, hardly tenable in view of the recent determination of the Landé factor by Kikoin and Goobar (1940), which shows that the super-currents are caused by movement of free electrons and are not merely simulated by electron spin. We stress this point since it will appear from our discussion that superconductivity is primarily an electric and not a magnetic phenomenon. The connection between electric and magnetic effects in superconductors has been expressed in rigorous form by the work of F. and H. London (1935). Their fundamental equation which connects the magnetic field with the current density does not, however, imply, as has sometimes been assumed (Burton *et al.*, 1940, p. 320), that the magnetic effects are the primary ones. The London theory is a phenomenological one which does not differentiate between cause and effect.

We have already given a short summary of the corresponding phenomena in superconductivity and liquid helium (Daunt and Mendelssohn, 1942), and it is not intended to discuss the analogy here in greater detail. Our hypothesis implies that the fundamental phenomenon in both cases is that of *frictionless transport*, and we shall now consider the experimental justification for this statement. In superconductors the evidence is provided by the classical current-

* Summaries of the phenomena in liquid helium have been given by Darrow (1940) and Jones (1940). A description of superconductivity has been given by Shoenberg (1938) and a summary of recent work by Jackson (1940)

potential measurement (figure 1 *a*) in which part of an electric circuit is superconductive. An ammeter in the circuit indicates a flow of electrons through the superconductor while a voltmeter shows zero potential between any two places of the superconductive part of the circuit.* Thus we have a transport of particles—electrons—through the superconductive metal in which no energy is dissipated. The corresponding effect in liquid helium II is illustrated by the transfer of helium atoms in a film over the wall of a beaker (Daunt and Mendelssohn, 1939 *a*) (figure 1 *b*). If the level inside the beaker is higher than outside, helium atoms are transferred out of the beaker until the difference in level has disappeared. The most remarkable fact about this transfer is that in contradistinction to the action of a syphon the rate of transfer is independent of the length of the path and of the difference in the height of the levels. This goes

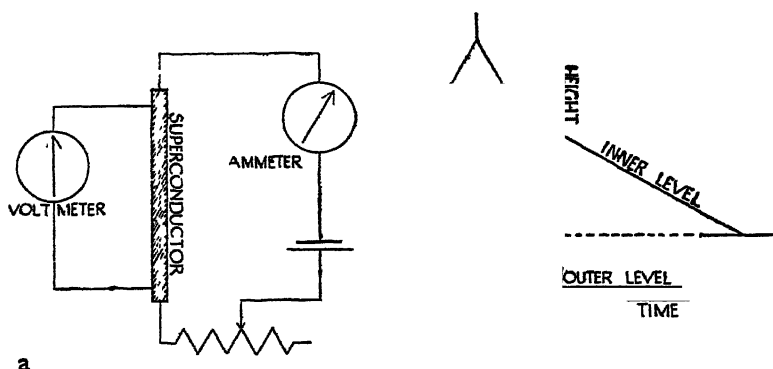


Figure 1. Experiments showing frictionless transport, (*a*) in a superconductor, and (*b*) in a film of liquid helium II. A sudden change in the level difference does *not* affect the rate of transfer.

so far that when the level difference is suddenly changed, no change at all is observed in the rate of transfer. Here again we have a transport of particles—in this case helium atoms—without friction, for friction would make the rate of transfer dependent both on the difference in potential and on the length of the path which the atoms have to traverse. The same phenomenon has been observed in flow of the bulk liquid through capillaries (Allen and Misener, 1939; Johns, Wilhelm and Smith, 1939) although not so clearly defined because it is partly masked by viscous flow of the liquid as a whole. However, in very fine capillaries ($\sim 10^{-5}$ cm.) the result is practically identical with the observations on films, and for wider capillaries it can be expressed in the form

$$r = R + n\Delta P, \quad \dots (1)$$

where r is the total rate of transport (volume/time), R the rate due to frictionless flow and $n\Delta P$ the amount transported by viscous flow under the pressure difference ΔP . The value of R and its dependence on temperature in the finest capillaries have been found to be of the same order as in films. This means that, as in superconductors, we are faced with the fact of frictionless transport through the rest of the substance. There is some evidence that both in superconductors and in liquid helium the frictionless transport shows a preference for the geometric

* The most accurate determinations (using persistent currents) show that the resistivity of a superconductor is less than 10^{-20} ohm-cm.

surface of the substance—a phenomenon which will be discussed later, and which does not affect our general argument.

The experiments thus show that there are “superconductive” electrons and “superfluid” helium atoms which move through their respective “container substances” without friction. When using such terminology we must, however, keep in mind that we cannot distinguish between individual electrons or individual helium atoms, and that these terms can only be used in allotting a certain number of particles to a given set of energy states. In the present case the particular set of states is distinguished by the fact that the electrons or helium atoms in this set—we will call them *z-particles* for short—do not dissipate momentum.

It must be pointed out straight away that the *z-particles* do not form a separate phase in either case. Apart from the lack of any direct experimental evidence to this effect, the thermodynamic functions, which we shall discuss later, show clearly that the appearance of the *z-particles* is not accompanied by a phase change in coordinate space. It has even been impossible to detect any increase in spatial order to account for the observed decrease in entropy which is linked with the

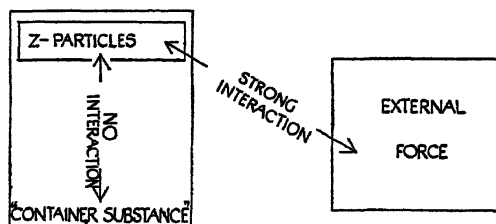


Figure 2. The peculiar position of the “aggregate” of *z-particles*; they exhibit no interaction with the rest of the substance, but strong interaction with an external force

appearance of the *z-particles*. There is, however, another way in which the assembly of *z-particles* is distinguished from any other assembly of spatially disordered, freely mobile particles belonging to a particular set of states. This is shown by figure 2, which illustrates diagrammatically the relation between the *z-particles*, the “container substance” and an external force (magnetic field, gravitation) acting on* the *z-particles*. There exists no interaction between the *z-particles* and the rest of the substance, but a strong interaction between them and the external force, which means that the *z-particles* react to it as a *separate aggregate*. More than this, the *z-particles* assume under the influence of such an external force the aspect of a spatially ordered aggregate whose degree of order can become manifest in macroscopic dimensions (e.g. the magnetic field of a persistent current). It seems therefore convenient to describe the *z-particles* as forming a particular state of aggregation—since we wish to avoid the term *phase*—which we will call the frictionless state of aggregation (*z-state*).

It must be stressed that this term does not of course refer to the superconductive metal or liquid helium as a whole but only to the assembly of *z-particles*, i.e. the “superconductive” electrons and the “superfluid” helium atoms. As the temperature of the substance is raised from absolute zero, the *z-state* weakens

* As will be seen in § 5 the conception of an “accelerating” force has to be qualified, since it seems possible that all frictionless transport is due to zero-point momentum, the acting force providing space order rather than acceleration,

and finally disappears. We will show in § 3 that this phenomenon can be described as a gradual decrease in the number of z-particles with increasing temperature. The frictionless state shrinks until, at the transition point or the λ -point respectively, the last z-particle disappears. In superconductors an indication of the change in population of the z-state is given by the threshold curve which measures the number of electrons that can be transported without friction in a given time over a superconductive surface of given dimensions at different temperatures. We have determined a similar quantity for the frictionless

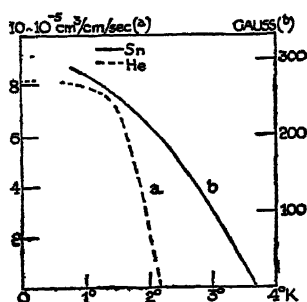


Figure 3. Critical transport rates (a) in liquid helium, and (b) in a superconductor (tin).

transport of helium atoms in films (Daunt and Mendelssohn, 1939 a). It is significant that this "rate of transfer"—the number of helium atoms transported without friction in unit time over a helium surface of given dimensions at different temperatures—was found to depend on temperature only. Figure 3 gives a typical threshold curve and the rate of transfer in comparison. These functions, which merely give the amount of z-particles collected in unit time, contain of course an unknown average velocity. If this velocity is independent of the temperature, as appears likely (see § 5), the curves can be regarded as indicating directly the relative number of z-particles as a function of the temperature.

§ 3. THERMAL ENERGY OF THE FRICTIONLESS STATE

The gradual disappearance of the frictionless state of aggregation is accompanied both in superconductors and in liquid helium II by an excess specific

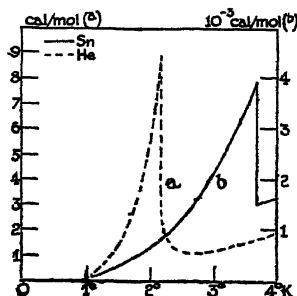


Figure 4. Specific heat (a) of liquid helium, and (b) of the free electrons in tin.

heat (figure 4). In superconductors it has been possible to separate the specific heat which is due to the electrons from that caused by the lattice vibrations (Daunt, Horseman and Mendelssohn, 1939). It has sometimes been assumed

that this excess specific heat, which is caused somehow by those particles participating in the anomalous state, must imply heat exchange between a system of these particles and the rest of the substance. The point has been discussed in the case of superconductors (Burton *et al.*, 1940, p. 298). The mechanism sketched in figure 2 on the other hand makes it very difficult to account for such an interaction. In fact, our definition of the *z*-particles—and this definition was derived from experimental results—would not allow direct thermal interaction with the “container substance” unless we provide for it by additional assumptions. Any such assumption, allowing energy exchange in one case and prohibiting it in another, would moreover be highly artificial.

This obvious discrepancy requires an explanation and it raises the question of the energy state of the *z*-particles, i.e. of the heat content of the *z*-state. The specific heats given in figure 4 refer of course to the contributions by the *z*-particles and by the normal particles, and it is clear that information on the thermal state of the *z*-particles by themselves must be obtained by experiments of a different character.

Such experiments have been carried out both for a superconductor and for liquid helium II. In the superconductor this is the determination of the Thomson heat (Daunt and Mendelssohn, 1938); the experiment and the result have been discussed in detail elsewhere (Daunt and Mendelssohn, 1945). The fact that the Thomson heat of a persistent current is zero means that the superconductive electrons do *not* exchange thermal energy with the rest of the substance. This leads to a peculiar consequence when we consider what must happen in a ring of superconductive metal with a persistent current running in it, which is warmed up from absolute zero. Since the electrons in the current are not in heat exchange with the lattice, they must remain energetically at absolute zero although the substance as a whole warms up.

An idea as to the thermal state of the superfluid helium atoms can be obtained from any experiment which causes their separation in space from the liquid as a whole. We have seen in the last section that, while there exists no space order generally, an external force has such an ordering influence (see figure 2). In determining the Thomson heat in a superconductor we have therefore made use

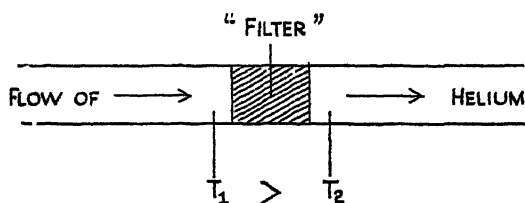


Figure 5 The mechano-caloric effect in which a separation of *z*-particles occurs. A flow of liquid helium II through a powder “filter” is accompanied with a heating of the liquid at the entrance and a cooling at the exit

of the influence of the magnetic field, and in the case of helium we can use gravitation as the acting force. In the experiment shown in figure 1*b* superfluid helium atoms were separated by surface flow. If the heat content of the superfluid helium atoms, like that of the superconductive electrons, is smaller than that

of the normal particles, the liquid drained off must be colder than that inside the beaker. We have carried out an experiment of this type (Daunt and Mendelssohn, 1939 b), using for practical reasons the flow of liquid helium II through tightly packed fine powder instead of flow over an exposed surface. A diagrammatical sketch of the method is given in figure 5. As was expected, a heating was observed at the entrance of the liquid into the "filter" and a cooling at the exit. H. London (1939) has since given a general thermodynamical interpretation of this mechano-caloric effect. Under the assumption of complete reversibility he obtains the following relation between the pressure (p) under which the liquid is forced through the power plug and the entropy difference (ΔS) between the helium flowing without friction and the bulk liquid :

$$\frac{dp}{dT} = \rho \Delta S, \quad \dots\dots (2)$$

where ρ is the density of the liquid. Our experiments (which had to be suspended in 1939) were of a preliminary character only, but the (dp/dT) -values obtained * yielded a large entropy change indicating that the superfluid helium atoms seemed to have lost *all* thermal energy.† Fortunately the experiment has since been repeated by Kapitza (1941), using a narrow gap instead of the powder plug. He could not only confirm the existence of the mechano-caloric effect and demonstrate with greater accuracy the complete loss of thermal energy suffered by the superfluid atoms, but by measuring independently the heat change connected with the effect, he has shown that the process is reversible.

It has to be admitted that we cannot yet establish the complete loss of thermal energy with the same accuracy for the superfluid helium atoms as in the case of the superconductive electrons.‡ Particularly the specific heat at low temperatures (Pickard and Simon, 1945) has been interpreted as corresponding to a "lattice entropy" (*cf.* H. London, 1939) while frictionless transport has been assumed at the same time. We will return to this question, which is not at all clear, in § 4. For the present we shall neglect this hypothetical residual entropy which, as the experiments show, does not interfere with frictionless transport.

The experimental result that in a "container substance" which has a finite temperature, the z-particles remain energetically at absolute zero but have momentum, leads to the simple but somewhat surprising conclusion that they must move without friction. For friction is the dissipation of momentum and any particle engaged in a process involving friction must therefore suffer thermal excitation. This means that the observation that z-particles remain thermally unexcited while being capable of translatory motion is in itself sufficient to

* The work will be published in detail later.

† Allen and Reekie (1939 a) have obtained (dp/dT) -values from the thermo-mechanical (fountain-) effect which is the thermodynamic reverse of the mechano-caloric effect. These values are of the same order as those from the mechano-caloric effect, but again not accurate enough to determine the exact entropy loss.

‡ Our conception of the properties of the z-state suggests that when a rotating mass of liquid helium is cooled below the λ -point, a frictionless current of z-particles will be set up which will persist when the momentum of the thermally excited particles has been dissipated. The decay of this persistent helium current will provide an equally sensitive measure for the frictional resistance of the superfluid atoms as its electric counterpart does for the resistances of the flow of superconductive electrons.

explain the phenomenon of frictionless transport. It seems permissible to speak of the entropy of the z -state since this state, as will be seen, forms a separate phase in momentum space, and it is clear that this entropy could not retain the value zero—as we know it does—when the state is agitated, unless this agitation does not involve friction.

The realization that the entropy of the frictionless state is zero at once removes the discrepancy with which we were faced when considering the excess specific heat of the z -particles. It also answers the question as to the significance of the entropy change accompanying the weakening of the frictionless state with increasing temperature. There exists no direct thermal interaction between the z -particles and the rest of the substance—the specific heat of the z -state is zero. The observed specific heat is clearly due to an excitation process in which z -particles are lifted into thermally excited states. When this occurs, they cease being z -particles and cease to be available for frictionless transport.

The anomaly in the specific heat connected with this change (figure 4) exhibits the familiar pattern associated with “cooperative” phenomena like ferromagnetism, rochelle-electricity and order-disorder transformations. The latter in particular have been made the subject of detailed statistical treatment, the methods of which have been applied to other phenomena, as for instance molecular rotation in solids. However, no general statistical treatment has yet been developed which can be rigorously applied to all these changes without regard to the particular physical process involved. Simon (1930) who first investigated these “quantum jumps” in solids has generalized them as excitations of internal

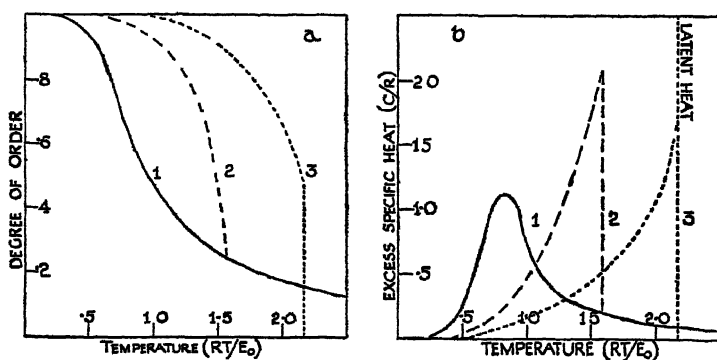


Figure 6. (a) Degree of order, and (b) excess specific heat of different alloys (theoretical) (1) Type AB with short range order, (2) type AB with long range order, and (3) type AB_3 with long range order. E_0 is the total energy of the transformation required to change the alloy from complete order to complete disorder (Cf Fowler (1936), Nix and Shockley (1938)).

degrees of freedom. The large amount of experimental data gathered in the last twenty years exhibits all stages of specific-heat anomalies ranging from the uninfluenced Schottky curve over λ -points to transitions involving a latent heat. This fact is important for our particular problem, since it shows that no clear-cut distinction is possible with regard to the mechanisms of the various excitation processes, and that to describe an unknown physical process on the strength of the observed specific-heat function as an order-disorder transformation

or a rotational change has little meaning. It usually means only that the particular statistical treatment can be applied, but it discloses nothing with regard to the *nature* of the process. In the case of the order-disorder transformations it has been possible to connect the theoretically determined states of order with specific-heat functions, which approach the observed ones, for a non-cooperative (short-range order, AB) type,* a cooperative (long-range order, AB) type, and a highly cooperative (AB_3) type. Figure 6 illustrates the point that the type of the change is not indicative of its nature. The degree of order—the inverse of the entropy—is a function, and in certain cases a direct measure, of the population of the lower state.

Considering the apparent lack of space order and the high symmetry of both the electron and the helium atom, we are forced to assume that the degrees of freedom lost by the assembly of z -particles are those of translatory energy. There is nothing unusual in the fact that at finite temperatures some particles in a substance are energetically at absolute zero. This is actually true for a proportion of the atoms in any solid at low enough temperatures. However, in the case of the z -state we deal with freely mobile particles. The nature of the phenomenon of frictionless transport is therefore the manifestation of an effect predicted by Nernst (1926) when he extended the third law to systems with freely mobile particles, i.e. gas degeneracy.

§4. THE ENERGY SPECTRUM

While thus frictionless transport is connected with gas degeneracy, this does not mean that it is a necessary consequence of it. In fact, the free electrons in the Sommerfeld model are in a highly degenerate state without exhibiting frictionless transport. The existence of z -particles, capable of translatory motion but retaining zero entropy, requires therefore not only degeneracy but degeneracy of a particular type. The particular qualification necessary for frictionless transport becomes apparent when we consider the term system of the Sommerfeld model (figure 7 *a*). A continuum of energy states is filled at absolute zero up to the level (L) which divides a completely filled set of states (A) from a completely unoccupied one (B). At finite temperatures an accelerated electron can and will dissipate energy, since it can be lifted to *any* place above (L). If we wish to allow for frictionless transport, we must prohibit states bordering (L), so that no dissipation of momentum can take place; i.e. we must introduce a *gap* in the term system (figure 7 *b*). As we have pointed out elsewhere (1945), in superconductors this gap forms the limit of the Fermi distribution at absolute zero. It follows from the considerations discussed in the preceding sections that such a term system in which the continua (A) and (B) are separated by a gap in the energy spectrum must be characteristic not only of a superconductor but of any other z -state. (The only other case known is, of course, liquid helium II.)

The existence of the gap signifies a separation in momentum space between the particles in (A) which cannot dissipate momentum and those in (B) which are thermally excited and acceleration of which results in friction since they can

* This function, a Schottky curve, has been observed by Simon (1926) in a number of pure metals, the nature of the change is unknown. Other examples of this type of anomaly are the paramagnetic salts and solid ortho-hydrogen.

be lifted to any higher state in the continuum (B). We can therefore speak of a "condensation" in momentum space, the z -particles forming the "condensed" phase. This phase forms the frictionless state of aggregation. Neither in superconductors nor in liquid helium is this a condensation to zero energy in the model sketched above. This means that neither case corresponds to a Bose-Einstein condensation since both have an appreciable zero-point energy. As mentioned in §3, it is by no means clear what happens to the "condensed" helium atoms in the liquid. The assumption that the rise in the total entropy of liquid helium below 1°K ., which Pickard and Simon found to be proportional to T^3 , is due to "lattice vibrations" can be brought in accordance, though with

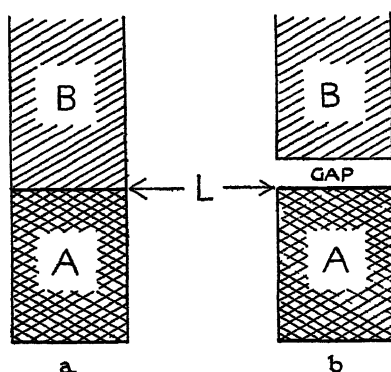


Figure 7 Term systems (a) of the electrons in a normal metal, and (b) as suggested for the electrons in a super-conductor or the atoms in liquid helium. A is the zero-point energy.

some difficulty only, with the phenomenon of frictionless transport, since it may be argued that the vibrations excited at these temperatures refer to large aggregates of atoms. On the other hand, the available data on the mechano-caloric effect seem to indicate that the *total* entropy disappears,* which would mean that we cannot separate the entropy of liquid helium into an anomalous and a normal part and that even at fairly low temperatures the overwhelming part of the specific heat is made up of the excitation of z -particles. More experimental data are clearly needed to settle this point. Apart from accurate determinations of the mechano-caloric effect, the necessary information can probably be obtained by measuring the coefficient of expansion below 1°K . The observed decrease in density with decreasing temperature is an anomalous effect due to "condensation" of z -particles and if at the lowest temperatures "lattice vibrations" are predominantly responsible for the entropy change, the coefficient of expansion should again change its sign.

The suggested term system removes the difficulty, which has sometimes been emphasized in the case of superconductors, that the z -particles appear to be not in *thermal* equilibrium with the rest of the substance. Such a state would of course not be stable. The reason why they remain energetically at absolute zero

* At 1.5°K , for instance, the observed difference between the entropy of the z -particles and the entropy of the whole liquid (ΔS in formula (2)) is $5 \times 10^{-2} \text{ cal/g}$., while the total entropy has been calculated (Simon, unpublished) as $5.04 \times 10^{-2} \text{ cal/g}$. The extrapolated "lattice entropy" ($\theta=15$, Pickard and Simon, unpublished) of $3.8 \times 10^{-2} \text{ cal/g}$ is too big to be accommodated in the difference between the first two figures.

even if the container substance has finite temperatures, is simply due to a lack of heat quanta large enough to lift z -particles over the gap. As the temperature of the substance is raised, the proportion of large quanta increases until the gap becomes insignificant. The shape of the specific heat curves (figure 4) indicates that the transition is not uninfluenced by the number of particles in the higher state, since it does not follow a Schottky function (figure 6*b*, 1), but indicates a cooperative phenomenon like ferromagnetism or the influence of long-range order on an order-disorder transformation. The rough term system given in figure 7*b* does not, of course, allow conclusions to be drawn as to the shape of the specific-heat curve, because it is completely empirical and no assumptions are made about the density of states above or below the gap. It must be emphasized that, since our considerations are based entirely on experimental observations, no theoretical explanation for the existence of such gaps in the energy spectrum can be given.

§ 5. ZERO - POINT DIFFUSION

So far our account has given no explanation of two outstanding characteristics of the z -state, the breakdown of frictionless transport when a critical rate of transport is exceeded and the high heat-conduction of liquid helium II. Both phenomena are, as we shall see, intimately connected. When investigating the high heat-conduction of the helium film (Daunt and Mendelssohn, 1939*a*), we were able to show that this also was due to a transport phenomenon, the flow

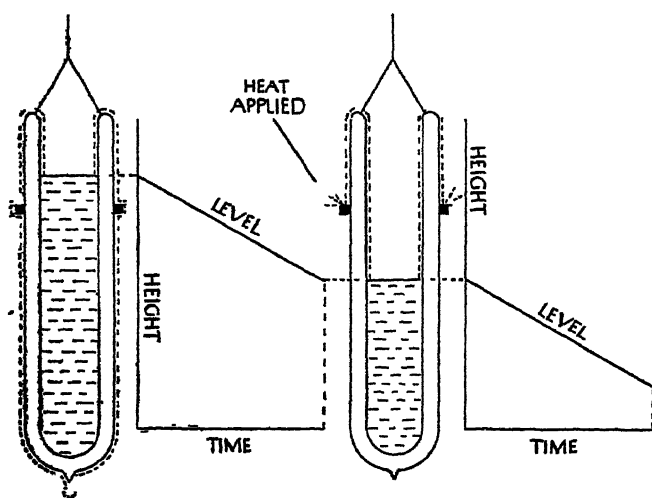


Figure 8. The rate of flow surface of liquid helium II is unaffected by the application of heat to the film. This can be explained by the assumption that the z -particles diffuse under their zero-point momentum.

of super-fluid helium atoms towards the place where heat was supplied. We then suggested that the heat conduction in the bulk liquid would probably be due to a similar though rather more complicated convection process. That this is actually the case has since been demonstrated by Kapitza (1940) in a number of experiments showing the existence of a counter current of liquid helium in a capillary under a temperature gradient. For an investigation of the mechanism of the process let us revert to the simple case of the helium film (figure 8).

We observed that the rate at which helium is transferred out of a Dewar vessel is *not* affected when heat is applied to the film outside the vessel. This not only demonstrates that the observed high heat-transport through the film is entirely due to a transport of helium atoms, but also that this process must evidently be *isothermal*, since a heating of the film at one place would have resulted in heat flow into the vessel by conduction. The salient features of the film transfer discussed above (§2) show that only z-particles are engaged in the movement and the heat supplied to the film must be taken up as heat of excitation which lifts the z-particles into the continuum of energy states and evaporates them. Heat is of course gained by the helium inside the vessel owing to loss of z-particles, but the *same* gain of heat occurs whether the film is heated or simply allowed to drop off. This explanation could be proved by an experiment (Daunt and Mendelssohn, 1939b) in which heat was supplied to the inside of a Dewar vessel where the concentration of z-particles was thus decreased (figure 9). In this

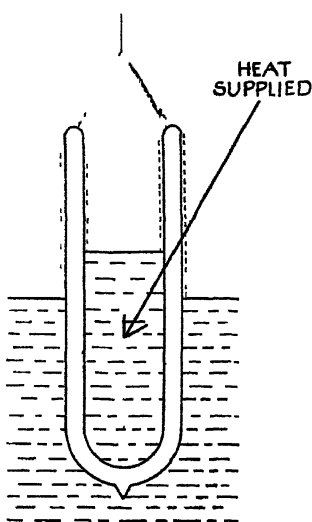


Figure 9 When heat is supplied to the inside of the vessel, the concentration of z-particles is decreased and zero-point diffusion takes place along the film from the bath into the vessel

case a transfer of z-particles *into* the vessel must take place, resulting, as was actually observed, in a rise of the liquid level inside the vessel. The observed effect clearly corresponds to the "fountain phenomenon" discovered by Allen and Jones (1938) in the bulk liquid. The comparatively simple conditions of our experiment now allow an interpretation of the whole set of effects connected with the heat transport in liquid helium II.

The important point in the experiment shown in figure 8 is that there is no "accelerating force" which drives the z-particles to the place at which heat is supplied. It is a striking feature of the experiments that, as was pointed out (Daunt and Mendelssohn, 1939a), a transfer always takes place to those parts where the film (i.e. z-particles,) is removed. Since the z-particles do not exchange energy with the rest of the substance, the force distributing them evenly over the available surface must be their own zero-point energy. The heat transport

in a film of liquid helium II is thus simply due to the *diffusion of z-particles under their zero-point momentum*.

In the bulk liquid, this zero-point diffusion will manifest itself as the tendency of the z-particles to take up a statistical distribution within the space occupied by the "container substance", and a macroscopic movement of z-particles must take place whenever this statistical distribution, i.e. the thermal equilibrium in the substance as a whole, is disturbed. It may be argued that such a "diffusion" process of a set of particles having zero entropy would be in contradiction to the third law. This is, however, not the case since the substance as a whole has finite entropy, and the experiments on helium II (Kurti and Simon, 1938) show that at very low temperatures, where the total entropy becomes small, the thermal changes associated with the flow phenomenon disappear. It is significant that the assemblies exhibiting the z-state—liquid helium and metal electrons—are distinguished by their high zero-point energy. The fundamental importance of the influence of zero-point energy on the phenomena in helium has been emphasized by Simon (1934), who pointed out that its high zero-point energy is responsible for the existence of *liquid* helium at absolute zero.

Our explanation of the heat transport in helium II is therefore the opposite to the one given by Tisza (1938) on the basis of the hypothesis that liquid helium is a Bose-Einstein gas. He attributes *no* momentum to the unexcited particles and explains the flow of heat as due to an osmotic diffusion of the excited particles in the bulk liquid under a temperature gradient. It is difficult to see how such a model in which the flow of unexcited particles only appears as a compensating displacement current can account for the fountain phenomenon. Quite apart from the fact that this model or a similar one proposed by F. London (1939) can never lead to *isothermal* heat transport, it is in direct contradiction to the experiment of figure 8, since there is no bulk liquid for the diffusion of excited particles. Such a diffusion of excited particles as visualized by Tisza may of course take place as a secondary process in the bulk liquid, and, particularly for cases of large heat flow with which the diffusion by zero-point energy cannot cope, become the preponderant feature.* This is probably the explanation for the peculiar negative fountain effect observed by Allan and Reekie (1939 b). The inversion point of the fountain observed by these authors will be due to the equilibrium between the two rates of diffusion. The striking high heat-conduction of the bulk liquid for small heat flow and in particular its dependence on the strength of the heat flow cannot however be explained by Tisza's model, while it agrees well with the process of z-particle diffusion outlined above.

The conception of zero-point diffusion of z-particles immediately leads to an explanation of the critical rate of frictionless transport. Taking again the case

* The establishment of a temperature difference in helium II, therefore, depends on two limiting factors, the rate of heat supply and the removal by osmotic and hydrostatic pressure of excited particles. The latter is no problem in our experiment (figure 8), where they are simply evaporated or drop off, but is important in the bulk liquid. The usual type of heat conduction experiments, employing capillaries closed at one or both ends, is thus bound to give results far too complex for interpretation, since it is quite impossible to separate the two limiting factors. A more thorough discussion of the heat-conduction process in helium II will be given by J. G. Daunt and the author later.

of the helium film, it is clear that no rise of temperature will take place as long as the z -particles which are excited by the heat supply can be replaced by diffusion. This rate is limited by the number of z -particles per unit volume (depending on temperature) and their zero-point velocity. Once this rate is exceeded, frictionless transport breaks down.

It seems reasonable to assume a similar process of zero-point diffusion for the superconductive electrons, since in superconductors, too, we are faced with a critical rate of frictionless transport, i.e. the threshold value. When applying these considerations to the superconductive electrons we arrive at an interesting conclusion. A current set up by zero-point diffusion of electrons must be essentially different from a normal current since the momentum of the flowing electrons is derived solely from collisions* as in the case of helium atoms. This means that the cause of a normal current, the accelerating force of the electric field, does not appear at all in the process. This peculiar phenomenon of a current in zero electric field is of course exactly what has been observed in superconductors (figure 1 *a*). According to our conception there should be another phenomenon, so far unobserved, by which a super-current is distinguished from a normal one. Since it is set up by collisions, the speed of propagation of a (heat) impulse should be different, the maximum speed of a signal depending on the zero-point velocity of the highest level in A (figure 7 *b*). The super-current will also be distinguished by having (in a conductor of uniform cross-section) uniform charge density.

Thus zero-point diffusion seems to explain the curious analogy from which our considerations originated (1942), i.e. that to the absence of potential differences in a superconductor there corresponds the absence of temperature differences in the helium film. The conception of zero-point diffusion also removes at once the difficulty of understanding the independence of the transport rate of z -particles of the *strength* of external forces, e.g. the pressure head in helium. It is clear that our statement (made in §2) concerning the action of an external force on the z -state has to be qualified. Indeed it seems doubtful whether in *any* of the transport phenomena a true particle acceleration occurs. It appears that the action of external forces on z -particles is rather that of creating a certain type of (space-) order than of accelerating particles.

To illustrate this point we consider an element (dx) of a long cylinder on the surface of which frictionless transport of z -particles can occur. This model corresponds to a superconductive wire or to a solid rod covered with the helium

* When we use here the ordinary kinetic picture of momentum transfer by collision, it is merely done to distinguish the process of zero-point diffusion from that of acceleration by a macroscopic field of force. It has, however, to be remembered that zero-point diffusion is bound to differ from an ordinary diffusion process by the fact that while momentum is exchanged, it is not dissipated, and it is doubtful whether the terms "mean free path" and "average velocity" can be used in their accepted meaning. Retaining such a kinetic interpretation of zero-point energy, it appears that in systems with high zero-point energy there must be a considerable dispersion of velocities. An indication of this seems to have been observed by Ganz (1940) when he found that the first indication of a heat impulse travelling through liquid helium was propagated with a speed of $\sim 10^4$ cm/sec, while the speed of the maximum of the impulse was only ~ 10 cm/sec. Incidentally, the secondary maximum observed by him may have been due to Tisza's diffusion of excited particles discussed above.

film. If at some place outside (dx), z -particles are removed, zero-point diffusion must take place and a macroscopic flow of z -particles in the x -direction will occur. This process entails no actual acceleration of z -particles but merely a greater mean free path in the x -direction. The flow of z -particles is accompanied in the case of electrons by the appearance of a magnetic (but not an electric) field and in the case of helium atoms by a moment of inertia. The states of flow and rest are thus energetically different but, as the experiments show, of equal (zero) entropy. If we make (dx) part of a closed ring, the state of flow, if once started, will persist since there are no means of energy dissipation and we arrive at a metastable state accompanied in superconductors by a persistent magnetic moment and in helium II by a persistent moment of inertia. The former phenomenon is well known, and it should be possible to detect the latter, particularly if these considerations can be extended to a tube filled with the bulk liquid.

Another way of describing this phenomenon is to say that the momentum of frictionless transport is not dissipated because it is zero-point energy.

§6 SURFACE TRANSPORT

In our 1942 paper, we have drawn attention to the peculiar tendency of the z -particles to move along the geometrical surface of the substance. As a tentative explanation it was suggested that this surface transport may be caused by the fact that the acting force only penetrates the substance to a certain depth and that below this depth no movement of z -particles will take place. It may also be that the disturbance of the structure of the "container substance" produced by the surface facilitates the thermal excitation of z -particles, a process about which nothing is known so far. In this case the depth of penetration should be to some extent independent of the particular type of disturbance and in some way be a measure of the limit of the free path of z -particles.

In superconductors the phenomenon of surface transport is a well-established fact, but in the case of liquid helium the evidence for it is not so strong. In view of the similarity of the transport phenomena in the film with those in the bulk liquid (Daunt and Mendelssohn, 1939 a ; Allen and Misener, 1939) it has been suggested that the film may extend below the liquid level and give rise to a transport of z -particles along the solid walls in the liquid. Subsequent experiments by J. G. Daunt and the author (1939 b), Allen and Reekie (1939 b), and Kapitza (1940) indeed seem to strengthen this view, but it must be pointed out that some of the results can be explained equally well by zero-point diffusion of z -particles in the bulk liquid. A surface flow of superfluid helium atoms which is analogous to the skin effect in superconductors does not however necessitate a higher concentration of z -particles near the walls as is postulated in the model suggested by F. London (1939). A decision as to whether and to what extent the frictionless transport in liquid helium is confined to the walls of the container has to be left to further experiments.

§7 BOSE-EINSTEIN CONDENSATION

The conception of the frictionless state of aggregation at which we have arrived in the preceding sections was reached merely by conclusions from the

existing experimental evidence and does not contain any theoretical hypothesis. Even so, a fairly consistent model could be obtained, the chief characteristics of which are those of a quantum liquid condensing in momentum space ; i.e. a set of freely mobile particles of zero thermal energy but an appreciable zero-point energy which is separated from the continuum of thermal states by an energy gap. In 1938 F. London attempted to explain the λ phenomenon of liquid helium in terms of the condensation of a Bose-Einstein gas, and it is interesting to see how far such a model can be used as a theoretical basis for the interpretation of our empirical conception of the z-state.

The chief difficulty is of course that Bose-Einstein statistics is not applicable to electrons in metals. However, as has been pointed out by Lindemann (1932), both Bose-Einstein and Fermi-Dirac statistics refer to idealized cases and it is quite possible that an intermediate treatment may prove equally applicable to electrons and helium atoms (cf. Daunt and Mendelssohn, 1945).

It is evident that the high zero-point energy of both liquid helium and the metal electrons plays a most important part in the behaviour of the z-state whereas the (ideal) Bose-Einstein gas has *no* zero-point energy. This is clearly the reason why Bose-Einstein statistics has failed to provide an explanation for the pressure-independent flow* and the isothermal heat transport in helium films. The process of zero-point diffusion which we have introduced in order to explain these phenomena does not exist in the ideal Bose-Einstein gas. Gogate and Rai (1944) have recently compared our rate of transfer of superfluid helium atoms in films (figure 3 *a*) with the velocity† of “non-energetic” particles derived from Bose-Einstein statistics and have found a fairly good agreement for the temperature dependency except for the lower temperatures. In their calculation they use the population of the lowest energy state as given by Bose-Einstein statistics to be

$$n' = n[1 - (T/T_0)^\sigma], \quad \dots\dots (3)$$

where n is the total number of particles per unit volume, T_0 the λ -point and σ has the value $3/2$. A much closer agreement can however be obtained, as we pointed out recently (1945), by using the value 5 for σ , and this is the value substituted by F. London (1939) in his modified theory of Bose-Einstein condensation. It may appear at first as if this remarkable agreement would provide a strong support for the application of Bose-Einstein statistics, but actually the contrary seems to be the case. The value ($\sigma=5$) was chosen by London not on any theoretical grounds, but merely as the value which happens to fit the observed

* Tisza has tried to explain this phenomenon by the accompanying thermal effects, while F. London postulates a dependency on the square root of the acting force. Both alternatives have been clearly disproved by the experiments (Daunt and Mendelssohn, 1939 *a*).

† In doing this, they tacitly assume the applicability of the quantum-mechanical formula $v \sim h/md$ for the zero-point velocity of mean free path, where m is the mass of the helium atom and d the mean free path. This is very tempting, particularly since adoption of the film thickness as value for d leads to a velocity value of the same order as the effective velocity observed in the film transport ($\sim 10^3$ cm./sec). The agreement is, however, deceptive, since the zero-point energy of liquid helium calculated by the same formula gives a value 10^4 times too small. We have therefore adopted the interpretation of the film thickness as a *limit* (corresponding to the lowest levels of A, figure 7*b*) of free path, given at the end of the first paragraph of §7 of this paper.

specific-heat curve of liquid helium.* As was shown in § 3, there exists a quite general statistical relation which is entirely independent of the particular type of statistics employed and which connects the population of the lower energy state with the excess specific heat. This relation simply states the fact that the population of the lower state determines the change of the degree of order and with it the excess specific heat in *any* thermal excitation of a degree of freedom. The case has been thoroughly discussed for order-disorder transformations and for molecular rotation (cf. Fowler, 1936). The agreement between the observed rate of transfer and equation (3) has therefore nothing to do with Bose-Einstein statistics, but merely re-states, and incidentally confirms, the experimental fact that the rate of transfer measures the population of the z -state in helium.

Thus while Bose-Einstein statistics fails to account for the excess specific heat of liquid helium and for the characteristic properties of flow and heat transport, the question arises whether it has any bearing whatsoever on the z -state as it is observed in liquid helium and superconductors. It seems that the position can be summarized as follows. The experiments show that both electrons and helium atoms can pass into a state in which they are capable of frictionless transport. These z -particles are energetically at absolute zero even at finite temperatures—they have zero entropy. The existence of particles in the lowest set of quantum states does not in itself lead to frictionless transport; it is necessary that these states should be separated from the rest of the energy spectrum by an energy gap. That means that a “condensation” in momentum space must take place if the frictionless state of aggregation is to appear, but this need not be a condensation into a state of no zero-point energy as in Bose-Einstein statistics. According to our considerations (§ 5) zero-point energy is necessary for the appearance of frictionless transport, and in fact no condensation to zero velocity takes place in superconductors and probably not in liquid helium either. The undisputed merit of Bose-Einstein statistics is that, while it cannot be rigorously applied to either superconductors or liquid helium, it is the only statistics so far developed which provides for a condensation phenomenon at all; and by introducing it F. London was the first to interpret the λ -phenomenon in terms of gas degeneracy.

The problem of liquid helium has been approached more recently by Landau (1941) who, instead of introducing a specific theoretical model, considers quite generally the quantization of the motion of liquids. He arrives, taking into account the experimental results, at two sets of hydrodynamic equations for the description of liquid helium. In its present state the theory appears still too general to present a rigorous model of liquid helium and it fails to account for the essential features of the transport phenomenon, as for instance isothermal heat transport, pressure-independent flow and the two components of the fountain effect. Its great advantage is that, in contra-distinction to the Bose-Einstein

* It is interesting that this purely empirical modification, introduced by London to allow for the “unideality” of his model, constitutes an approach to our suggested term system (figure 7 *b*). It has been shown that the adoption of ($\sigma=5$) instead of ($\sigma=3/2$) seems to produce the postulated gap in the energy spectrum of helium between the set of ground states and the thermally excited states.

model, it does not require modification of an already existing framework but seems to be capable of application to the case under consideration by mere specialization.

§ 8 CONCLUSION

The picture of the z -state derived from the experimental facts is far too consistent to permit us to dismiss the similarity between superconductivity and liquid helium II as merely accidental or superficial. Summarizing our conclusions, we arrive at three salient features of the frictionless state of aggregation : (a) The z -particles are energetically at absolute zero. (b) The zero-point energy is divided by an energy gap from the thermally excited states. (c) The transport phenomena are caused by a diffusion of z -particles under their zero-point momentum

(a) is fairly well established by direct experiment. The term system suggested by us to account for (b) is a very rough model and we can hardly expect it to yield more than a qualitative explanation, particularly as all evidence for the existence of the gap is indirect. (c) is of course closely connected with (a) and the fact that frictionless transport is only observed in certain metals is explained by (b). (c) is strongly supported by all the available experimental evidence and appears to provide a very satisfactory explanation for the transport phenomena in superconductors and liquid helium. The conception of zero-point diffusion would seem to be the most important conclusion reached in our considerations.

The foremost application of these considerations will be to provide a working hypothesis for experimental research in a field where so far nearly all important progress has been achieved by accidental observations. It is clear that our model can be tested by a great variety of experiments, a number of which has been indicated in this paper. Apart from the application to experimental work it is hoped that the above considerations will also aid the development of a theory of superconductivity and liquid helium. The most serious gap in this respect is the lack of a kinetic interpretation of a system of freely mobile particles with no thermal but finite zero-point energy.

Finally one may ask whether the z -state is realized anywhere in the world except in the laboratory. The high densities in stars and possibly in some of the planets may provide suitable conditions for its appearance in macroscopic dimensions. Such a high concentration of particles also occurs in the nuclei of heavy atoms which have been described by Bohr and Kalckar (1937) as droplets of quantum liquid, and Teller and Wheeler (1938) have drawn attention to a resemblance between the term system of liquid helium and that of the nucleus.

§ 9 ACKNOWLEDGMENTS

In the past years the author had occasion to discuss various aspects of the problem with a number of colleagues, and he would like to take this opportunity

of expressing his gratitude to them. He is particularly indebted to Dr. J. G. Daunt, who has helped greatly in the clarification of many points, and to Dr. F. Simon, F.R.S., for providing him with numerical data from unpublished work and for much valuable criticism.

REFERENCES

- ALLEN and JONES, 1938. *Nature, Lond*, **141**, 243
- ALLEN and MISENER, 1939. *Proc Roy Soc., A*, **127**, 467
- ALLEN and REEKIE, 1939 a. *Proc Camb. Phil Soc* **35**, 114
- ALLEN and REEKIE, 1939 b. *Nature, Lond*, **144**, 475
- BOHR and KALCKAR, 1937. *Kgl. Danske Videnskab Selskab*, **14**, no. 10
- BURTON, SMITH and WILHELM, 1940. *Phenomena at the Temperature of Liquid Helium* (New York: Reinold)
- DARROW, 1940. *Rev. Mod Phys.* **12**, 257
- DAUNT, HORSEMAN and MENDELSSOHN, 1939. *Phil Mag* **27**, 754
- DAUNT and MENDELSSOHN, 1938. *Nature, Lond*, **141**, 116, 1939 a. *Proc Roy Soc A*, **170**, 423, 1939 b. *Nature, Lond.*, **143**, 719; 1942. *Nature, Lond.*, **150**, 604, 1945. *Proc. Roy Soc A* (in the press)
- FOWLER, 1936. *Statistical Mechanics*, 2nd ed. (Cambridge)
- GANZ, 1940. *Proc. Camb. Phil. Soc.* **36**, 127.
- GOGATE and RAI, 1944. *Nature, Lond.*, **153**, 342
- JACKSON, 1940. *Rep. Progr Phys* **6**, 335.
- JOHNS, WILHELM and SMITH, 1939. *Canad J Res.* **17**, 149.
- JONES, 1940. *Rep. Progr. Phys.* **6**, 280
- KAPITZA, 1940. *J Phys U.S.S.R.* **4**, 181; 1941. *Phys Rev.* **60**, 354.
- KIKOIN and GOOBAR, 1940. *J Phys. U.S.S.R.* **3**, 333.
- KURTI and SIMON, 1938. *Nature, Lond* **142**, 207
- LANDAU, 1941. *Phys. Rev.* **60**, 358
- LINDEMANN, 1932. *Quantum Theory*, vii (Oxford).
- LONDON, F., 1938. *Nature, Lond*, **141**, 643; 1939. *J Phys. Chem* **43**, 49.
- LONDON, F. and LONDON H., 1935. *Proc Roy. Soc. A*, **149**, 71.
- LONDON, H., 1939. *Proc. Roy. Soc. A*, **171**, 484.
- NERNST, 1926. *The New Heat Theorem* (London: Methuen).
- NIX and SHOCKLEY, 1938. *Rev. Mod Phys* **10**, 1
- PICKARD and SIMON, 1945. *Proc. Roy Soc. A* (to be published shortly).
- SHOENBERG, 1938. *Superconductivity* (Cambridge).
- SIMON, 1926. *Ber Akad. Wiss Berlin*, **33**, 477; 1930. *Erg Exakt. Naturw.* **9**, 222; 1934. *Nature, Lond*, **133**, 529.
- TELLER and WHEELER, 1938. *Phys Rev.* **53**, 778.
- TISZA, 1938. *C.R Acad Sci., Paris*, **207**, 1035, 1186.

A DOUBLE-REFRACTION METHOD OF DETECTING TURBULENCE IN LIQUIDS

BY A. M. BINNIE

Communicated by E. J. Bowen, F.R.S. ; MS. received 17 March 1945

ABSTRACT The doubly-refractive properties of a weak solution of benzopurpurin were used in a study of the onset of turbulence when the liquid flowed through a long horizontal glass pipe. A beam of polarized light was arranged to traverse a diameter of the pipe remote from the ends, the nicols being set to give extinction when the liquid was stationary. The emergent light fell on a photo-cell, which was connected through an amplifier to a cathode-ray tube.

With unidirectional flow, photographs of the trace on the screen showed a steady straight line when the velocity was small. Disturbances appeared at a Reynolds number R of about 1770, and complete turbulence was established when R attained a value of 2900.

With forced oscillating flow, the trace at low velocities was a sinusoidal curve, on which, at a critical Reynolds number, small superposed ripples were observed. Even when the motion was greatly increased the intricate traces recurred perfectly, but this phase finally broke up into complete turbulence as soon as the amplitude of the surface movement in the bottles at the ends of the pipe reached a certain limit.

§1 INTRODUCTION

CERTAIN colloidal solutions exhibit the property of streaming double refraction. Most of the quantitative experiments which have been carried out on this phenomenon have been concerned with the problem of the shapes of the particles in solution, and they have been performed in the well-known concentric cylinder apparatus. More recently, however, this property has been used by Alcock and Sadron (1934) to provide a means of measuring the velocity distribution in various kinds of laminar flow. The properties of many liquids, suitable for the purpose, have been examined by Weller, Middlehurst and Steiner (1942), who gave numerous references to previous work. Andrade and Lewis (1926), who used rotating cylinders, remarked that the onset of turbulence was favoured by the presence of colloidal particles, but apart from a brief note by Hauser and Dewey (1939), no trace has been found of any attempt to use the method deliberately for detecting turbulence. However, the underlying principle appeared promising for the purpose and worth detailed attention. Accordingly both unidirectional and forced oscillating flow in a circular pipe were examined by this means at the Engineering Laboratory, Oxford. Many investigators* agree that, in the former, laminar flow breaks down at a Reynolds number of about 2100, calculated with the pipe diameter and with the mean velocity of flow. On the other hand, no previous information is forthcoming concerning forced oscillating flow.

* Cf. Goldstein (1938).

§2. THE WORKING LIQUID

The best known colloidal solutions possessing the required properties are vanadium pentoxide (V_2O_5), the red dye benzopurpurin 4B and certain clay suspensions. The first two contain filamentous particles; when stationary, the filaments are oriented at random owing to thermal agitation, but when in slow motion they cause the liquid to exhibit doubly refractive properties if viewed at right angles to the direction of flow. It appears from a review by Edsall (1942) of the present state of knowledge that there is a general measure of agreement amongst physical chemists concerning the nature of this effect. In steady laminar flow, the filaments as they are carried with the stream rotate about axes perpendicular to the direction of streaming, but they do so with an angular velocity which is a minimum when they lie along the stream. Thus at any instant there will be more particles in this position than in any other. This explanation also shows why these solutions, unless highly diluted, exhibit non-Newtonian effects, such as the distortion, examined by Lawrence (1935), of the parabolic velocity distribution in laminar unidirectional flow in a pipe. For it is clear that the filaments, as they whirl along, must be comparatively far apart if no interaction is to occur.

Thus for the present purpose it was necessary to select a solution sufficiently weak to show no appreciable anomalous properties, yet strong enough to exhibit measurable optical effects. After some preliminary tests, 0.25% benzopurpurin solution was chosen because it combines ease of preparation in quantity with a comparatively powerful double-refraction effect. The solution was prepared 2 l. at a time; 0.5 gm. of the powder was ground up in a mortar with a little distilled water, and the volume was made up to 0.5 l. After a slight warming the powder was completely dissolved, giving a clear red solution which was then diluted to 2 l. The alternative procedure of adding the powder to the full volume of water was found to be ineffective; the powder then refused to dissolve even on boiling. A solution of 80 gm. of potassium sulphate in 900 cc. of distilled water was then made, and 200 cc. of this was added to each 2 l. of benzopurpurin to form the working liquid. The salt solution causes the dye molecules to aggregate to particles of a size suitable for the double-refraction phenomenon, but larger quantities throw out the dye as a flocculent precipitate. The aggregation is caused by the electrical effects of the monovalent kation; the presence of kations of higher valency, derived, for example, by corrosion, must be guarded against in view of their powerful coagulating effect. Certain parts of the apparatus had to be made of metal, therefore the sulphate was used in preference to the more corrosive chloride.

One disadvantage of benzopurpurin as a working liquid is the change of its properties with time. For a few hours after mixing it was found to be inactive. On the first day after mixing its optical power was sufficiently strong for the experiments to proceed, but this increased too much during the day's work for the results to be entirely satisfactory. To avoid this difficulty, observations were then confined to the third and to the seventh days, over which the rate of development was slow. The growth of the particles depends upon the temperature, but this could not be kept constant; hence the optical properties of

the liquid on the two test days was not the same in successive sets of experiments. The liquid remained active for periods longer than a week, but it gradually became cloudy and therefore unsuitable.

When it was desired to interpret the observations in terms of Reynolds numbers, considerable difficulty was experienced because the kinematic viscosity of the working fluid was not free from uncertainty. Mark (1940), who described a benzopurpurin solution as "elastic", explained that its viscosity is maximum at very low shearing stresses, diminishing in an exponential manner to a steady Newtonian value as the shearing stress is increased. These effects were discovered in the working fluid, although it was so dilute. Experiments were made with a viscometer of the Couette type, in which the diameter of the outer rotating cylindrical surface was $1\frac{3}{8}$ in. and that of the inner cylinder (hung on a torsion wire) was $\frac{5}{8}$ in. At a mean velocity gradient of 3 sec.^{-1} the viscosity of the three-day-old fluid was found to be 36% in excess of that of water, while at 9 sec.^{-1} , the limit of the apparatus, the corresponding value fell to 21%. At a very high velocity gradient, estimated at 800 sec.^{-1} , observations with an Ostwald (capillary) viscometer gave the excess of kinematic viscosity as only 3%. The experiments to be described lay in a comparatively small intermediate range, and the onset of turbulence in unidirectional flow was at a mean velocity gradient of about $17\frac{1}{2} \text{ sec.}^{-1}$. The density of the working fluid was $\frac{1}{2}\%$ greater than that of water. On the assumption that the excess of kinematic viscosity diminished exponentially with the mean velocity gradient, logarithmic plotting of the above results showed that at $17\frac{1}{2} \text{ sec.}^{-1}$ the excess was about 9%, and, as no sensible change could be detected after seven days, this figure has been used in all the calculations.

It should be borne in mind that the viscosity of these anomalous solutions is known to depend partly on the previous mechanical history of the solution; vigorous treatment breaks up temporary linkages of the particles and for a time reduces the viscosity. Now in the unidirectional flow experiments the working liquid passed continuously through a centrifugal pump, and the oscillating flow apparatus was operated as violently as possible for a few minutes at the commencement of each series of tests. Therefore viscosities deduced from measurements made under the quiet conditions in the Couette viscometer were probably somewhat higher than those obtaining in the experiments. Support for this view is provided by the unidirectional flow tests described in §4. There it was found that the critical velocity occurred at a Reynolds number of about 1970, which is slightly lower than that established by other methods.

§3. THE OPTICAL APPARATUS

The horizontal glass pipe, through which the liquid passed, was 15 ft. long. It was made up of three 5-ft. lengths accurately fused together, the mean internal diameter of the central length being 0.94₃ in. At the central section, distant 95 diameters from the ends, the flow was examined by means of a photo-electric cell. The source of light was a 125 w. tungsten arc lamp (pointolite), the horizontal beam from which passed successively through a convex lens (to make the beam parallel), a cell containing a solution of ferrous and copper sulphates (to absorb injurious heat rays), an adjustable stop set at about $\frac{7}{32}$ in. diameter,

the first nicol or polarizer, a diameter of the pipe, and finally the second nicol or analyser. The axes of the nicols were fixed at 90° to each other, both being at 45° to the axis of the pipe. In order to preserve the approximately parallel nature of the beam, a rectangular glass cell full of distilled water was fitted round the section of the pipe which was under observation.

When the liquid was stationary, with the dye filaments oriented at random, the only light which emerged was that due to lack of parallelism in the beam, causing unequal reflection losses of the components of the plane-polarized light along and across the tube, with consequent imperfect extinction by the nicols. But, on putting the liquid into motion, the double-refraction effect caused the emergent beam to increase in strength. It fell on a photo-electric electron multiplier, consisting of a caesium photocell whose primary emission was magnified about 2.5×10^4 by seven stages of secondary photo-emission controlled by crossed electromagnetic and electrostatic fields. It is free from inertia effects in rapidly varying light. The output of the multiplier was led through a single stage of valve amplification, having a voltage gain of about 20, to the Y plates of a gas-focused cathode ray tube, the screen of which had a long after-glow. Thus, even at the very low velocities of sweep which were used, it was possible to compare the traces of successive cycles. The time base was of orthodox type, a condenser being charged through a pentode valve and discharged by contacts mechanically operated by the piston rod of the reciprocating pump described in §5. In order to secure the necessary stability, all potentials were supplied by dry batteries and accumulators. With a Leica camera, 56 photographs of unidirectional flow were obtained and 123 of oscillating flow, those reproduced in figures 2, 5 and 6 were taken at approximately half gain, which was found to be the most suitable ratio because at full gain, in spite of many precautions, the trace was disturbed by stray fields. The zero level of the trace was indicated by projecting the images of spots or arrows on to the screen.

In order that a satisfactory comparison might be made between the two types of flow, no adjustment was made to the glass tube or to the optical apparatus in the course of the work detailed in the following pages. Owing to the difficulty of obtaining a circulating pump, the experiments on oscillating flow were performed first. Here, however, the work will be described in the reverse but logical order.

§4 THE BREAKDOWN OF UNIDIRECTIONAL LAMINAR FLOW

As shown diagrammatically in figure 1, the inlet end of the glass pipe was joined by hose and a short length of 1-in. stainless-steel pipe to an aspirator bottle A. The outlet end was similarly connected to glass tubing leading to the large bottle B; this served as a reservoir from which the pump C drew its supply. The pump was of the centrifugal type, driven by an electric motor; its interior was chromium-plated and, to give additional protection, the plating was coated with bakelite varnish. The delivery from the pump passed either to the bottle A through glass tubing and an 11-ft. length of 1-in. stainless-steel pipe or back to the bottle B through a by-pass. Thus the flow through the glass pipe was turbulent at entry; it was effected by gravity only, and any pulsations developed in the pump could not reach the section under observation. The flow was con-

trolled by means of pinch-cocks on the by-pass and near the outlet end of the glass pipe, and also by altering the speed of the pump. No difficulty was experienced in maintaining constant conditions. The apparatus required a charge of about 8 l. of working fluid.

To measure the flow, a stainless-steel orifice-plate was inserted at D, and open glass stand-pipes fixed to a vertical scale were connected to both sides of it. In view of the limited supply of working fluid, it was thought preferable to use tap-water when determining the coefficients of discharge of the orifice at various rates of flow. In order that the observers should not be biased in any way, this calibration was not done until all the photographs had been obtained.

Observation revealed, and the camera recorded (figure 2), that at low velocities the trace on the screen was a horizontal straight line, which moved upwards further from the zero level when the velocity was increased. The zero with the

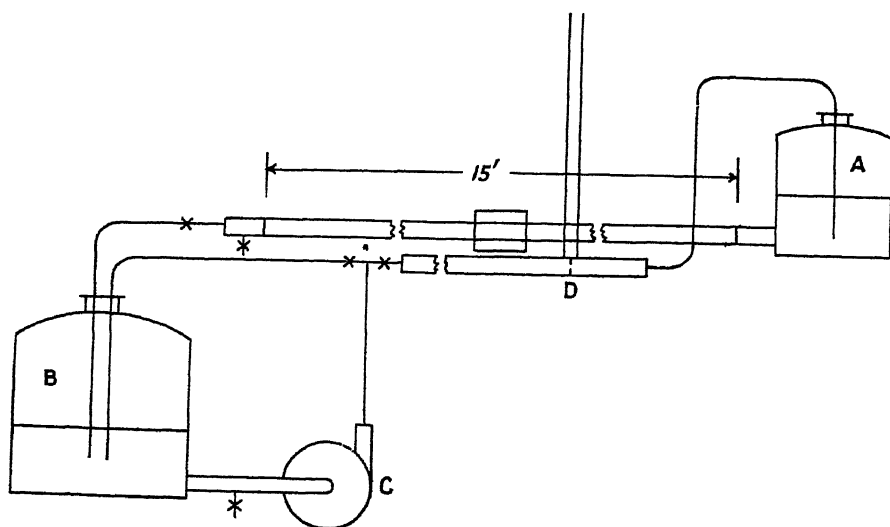
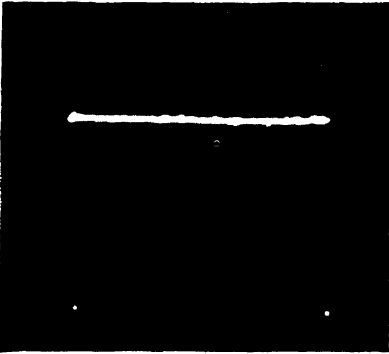


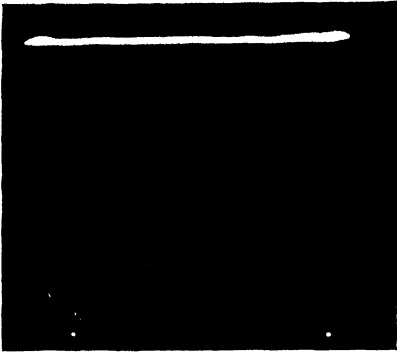
Figure 1 Elevation of the unidirectional flow apparatus

tungsten arc on and the liquid stationary is indicated on the photographs by two spots; when the light was turned off, the trace descended only by an amount approximately equal to its own breadth. The sweep was set to operate at the constant speed of 20 c/min. At higher velocities of flow, momentary oscillations occurred, evidently due to the "flashes" described by Reynolds. As the velocity was still further increased, the disturbances became more numerous and the frequency of the ripples went up. Finally a velocity was obtained when the trace was always in a state of agitation.

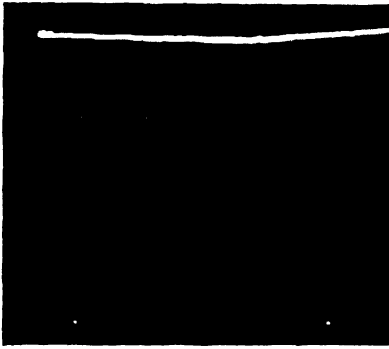
With the three-day-old liquid, photographs were secured of complete steadiness at a Reynolds number $R=1970$, and of disturbances at $R=2270$. But a closer "bracket" was obtained after seven days, therefore a selection of photographs from this test is reproduced in figure 2, where (i) and (ii) show the steady straight lines at Reynolds numbers of 750 and 1900. At $R=1970$ there were two "flashes" in five minutes, but attempts to photograph them were unsuccessful; however, a picture (iii) taken during the intervening periods shows that the steady



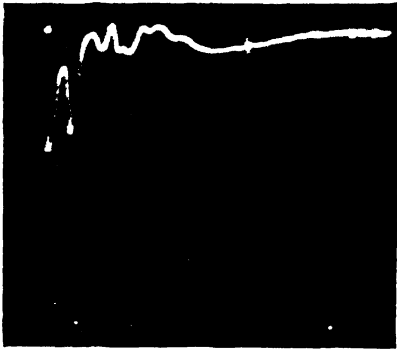
(i) $R=750$



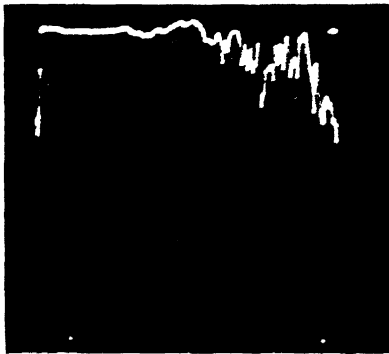
(ii) $R=1900$



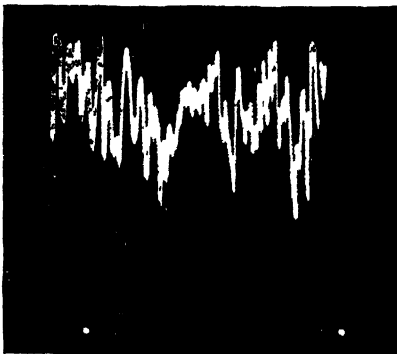
(iii) $R=1970$



(iv) $R=2030$



(v) $R=2620$



(vi) $R=2910$

Figure 2. Breakdown of laminar unidirectional flow.

line was slightly bowed. At $R=2030$ the disturbances were marked, a period of rippling and its termination being shown on (iv). At higher velocities the quiet periods were shorter and the rippling more violent, although even at $R=2620$ (shown in v) the bursts of turbulence did not often last longer than two cycles. In this phase the trace in an inactive period was seldom perfectly horizontal, and the approach of a turbulent spell could often be detected by a gentle fall of the line. At $R=2910$ (vi) really quiet intervals never occurred, and turbulence was completely established, while at still higher velocities the amplitude and frequency of the ripples further increased.

These results show at once that the apparatus gives much the same value of the critical Reynolds number as that obtained by earlier methods, but further information may be obtained from the photographs. Figure 3 gives the distance of the

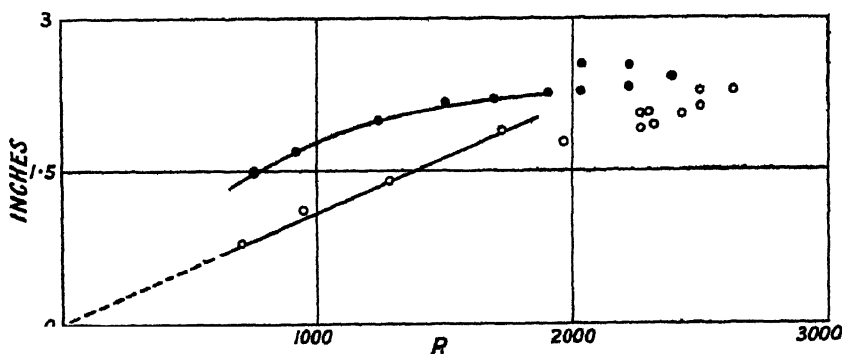


Figure 3. Deflexion of trace in unidirectional flow.

○ After 3 days. ● After 7 days.

straight traces from the zero with no flow plotted against R . Little reliance can be placed on points beyond the critical value because, as explained above, the line in a quiet period was not often truly horizontal, and these measurements are therefore dubious. Confining attention to laminar flow, we see that in the three-day tests the deflexion was linearly dependent on R , and this suggests little orientation of the particles at the very low tangential stresses set up in such a wide pipe. Had the orientation been complete, the curve would have become horizontal; in the seven-day tests it did in fact tend to do so, but the onset of turbulence cut off the finish of the process. These ideas are confirmed by the observation that, when the flow was stopped, the deflexion of the trace did not immediately fall to zero, from which it follows that the particles cannot have been very small. Thus the low degree of orientation must be attributed either to lack of anisometry in the particles, which seems improbable in view of their known filamentous shape, or to the small tangential stresses.

The investigation having for its main object the detection of turbulence, no attempt was made to analyse the turbulent vibrations, but it may be remarked that the laminar flow evidently broke up into vortices not much smaller than the cross-section of the light beam used. Only if the vortices were considerably less than this would the double refraction be reduced to zero, and this was not observed.

§5 THE BREAKDOWN OF OSCILLATING LAMINAR FLOW

It is desirable first to examine theoretically the nature of forced oscillating laminar flow in a pipe—a problem which has been attacked by Southwell and his collaborators, e.g. Christopherson, Gemant, Hogg and Southwell (1938). They considered a U-tube (composed of a horizontal pipe connected at each end to a bottle) in which the motion was maintained by a pulsating air pressure. Being concerned principally with establishing the basis of their method of viscometry, they concentrated their attention on the relation between the amplitudes of the applied pressure and of the surface movement in the bottles. Here, however, we are interested in the distribution of horizontal displacements over a cross-section of the pipe which results from a measured amplitude of the surface movement in the bottles. Accordingly Southwell's analysis has been extended. The details are relegated to the Appendix, but the results are shown in figure 4. This is a vector diagram drawn for the motion

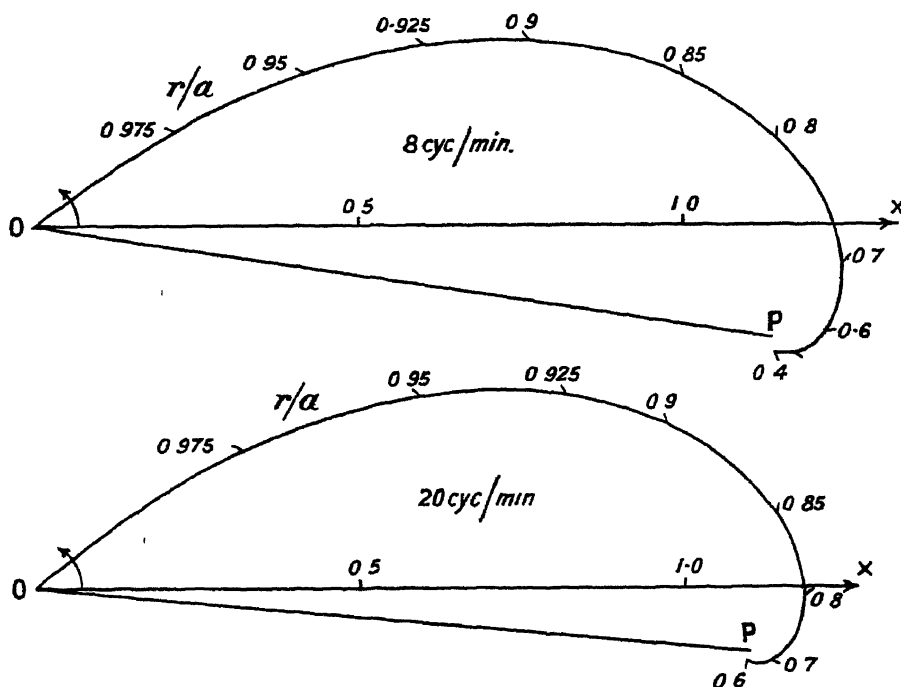


Figure 4 Theoretical distribution of displacement in laminar oscillating flow.

in the 15-ft. pipe of water at 15° C. at 8 and 20 c./min., these frequencies being the slowest and fastest used in the experiments. The direction of the line Ox represents the amplitude in the bottles. The curve is graduated in terms of the ratio of the radius r under consideration to the radius a of the pipe, and the magnitude and phase angle of the horizontal displacement are given by the straight line drawn from O to the curve. The distribution of these displacements has a kind of inverse analogy to the well-known skin effect which occurs when high-frequency alternating current passes through a solid circular conductor. It will be seen that at 8 c./min. the liquid within a radius of about $0.4a$ oscillates with

little relative movement. At 20 c./min. this radius is increased to about $0.6 a$; it is only at frequencies beyond the scope of the experiments that almost the whole of the liquid oscillates virtually as a solid plug with large shearing stresses close to the wall. It should be noted that the ratio of the horizontal displacements in the pipe to the amplitude in the bottles is independent of the latter when the other variables are unchanged.

In the first set of experiments the glass pipe was connected at each end to an aspirator bottle of diameter 5.6 in. In one bottle, off which the top was cut, the movement was measured with a stainless-steel hook gauge, surface disturbances being eliminated by perforated bakelite baffles. The other bottle was joined by rubber tubing to a single-cylinder reciprocating pump of diameter 3 in. and stroke $1\frac{5}{8}$ in., from which the valves were removed. The piston was actuated by a Scotch crank driven through reduction gearing by an electric motor, so that its motion was truly simple harmonic. It was found that if, on starting, the rubber tubing was connected when the piston was at the middle of its stroke, no creep in the excursions of the liquid took place. The amplitude of the motion of the liquid was controlled with four bottles (serving as capacities), joined singly to the connexion between the pump and the aspirator bottle; they could be turned on or off at will, and the effective volume of one could be finely adjusted by partially filling it with water. To satisfy the requirement that the pressure imposed on the working liquid should vary in a sinusoidal manner, the pressure changes in the air were kept small and nearly isothermal by running the pump only at very low speeds. The apparatus required about 5 l. of working liquid.

The method of test was to take a series of photographs at constant pump speed, the amplitude of the movement in the aspirator bottles being varied by means of the capacities. The surprising fact at once emerged that over a considerable range the trace at each amplitude, although considerably agitated, recurred exactly for an indefinite length of time. The break-up of this phase at 13 c./min. is illustrated by the series shown in figure 5, in which are indicated the travels measured in the aspirator bottle and the Reynolds numbers. The latter will now be formed with the maximum velocity of the motion, with the simplifying assumption that the velocity was uniform over the cross-section. The zero at no flow is here indicated by the tips of the arrows. The first photograph shows recurring flow at a travel of 0.55 cm.; the ends of the stroke correspond to the undisturbed portions of the trace, which, as in the unidirectional experiments, did not instantly descend to zero when the velocity became very small. At 0.80 cm., (ii), the trace still recurred. The next photograph records two consecutive cycles, the only difference which can be detected between them being the short-circuit of the prominent inverted W on the right. Again, the ends of the stroke may be distinguished by the quieter motion there, which, being slower, caused a thickening of the trace. Finally, at 1.61 cm., (iv), the diagrams did not recur, although the differences between them was not marked, but at still greater travels the motion of the trace was wildly irregular. These results indicate that under certain conditions a recurring system of vortices was formed, which was destroyed when the motion became too violent. The available information

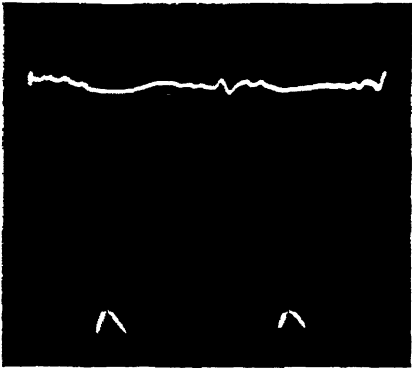
concerning the region of breakdown is summarized in table 1, in which the results of the one-day test (mentioned in § 2) have been included.

Table 1. The breakdown of recurring turbulent flow

Age of fluid	c./min	Travel (cm.)	<i>R</i>	
One day	20	0.97	6360	Recurring
"	"	1.30	8480	Almost recurring
"	"	1.41	9220	Not recurring
"	16½	1.24	6840	Recurring
"	"	1.52	8360	Not recurring
"	13	1.03	4610	Recurring
"	"	1.39	5880	Almost recurring
"	"	1.51	6750	Not recurring
Three days	20	0.98	6500	Recurring
"	"	1.32	8800	Not recurring
"	13	1.08	4840	Recurring
"	"	1.39	6190	Not recurring
Seven days	20	0.95	5810	Recurring
"	"	1.32	8050	Not recurring
"	13	0.80	3480	Recurring
"	"	1.10	4790	Almost recurring
"	"	1.25	5460	Almost recurring
"	"	1.61	6990	Not recurring

No correlation can be discovered among the values of *R*, but there are indications that it was the travel which was the determining factor. Allowing for the difficulty of estimating when recurrence was perfect, it may be suggested that a travel of about 1.25 cm. was the criterion of breakdown.

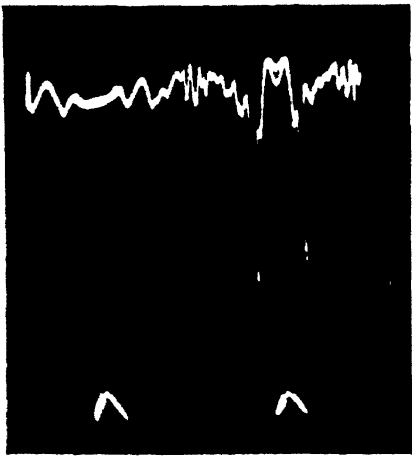
Attention was then concentrated on the first signs of turbulence. With a fresh charge of liquid, the trace at very slow flows was seen to be a smooth sinusoidal curve indicating laminar conditions. As the amplitude was increased, small ripples appeared at first, but later the harmonic nature of the picture was broken up into the form indicated in figure 5 (1). These experiments were found difficult and unsatisfactory because under some conditions the travel did not remain steady. Consequently suspicion fell on the apparatus used to maintain the motion. It has been pointed out by den Hartog (1940) that, if a forced oscillation is maintained with a spring of varying stiffness as coupling, the resulting motion may be unstable. In the experiments, the damping was considerable owing to the destruction of kinetic energy when the liquid entered an aspirator bottle; free oscillations were in fact nearly dead-beat. Thus the power required to drive the apparatus, and consequently the compression and varying elasticity of the intermediate air, seemed unnecessarily large. Moreover, at 20 c./min. the travel required for laminar flow was too small to be measurable with reasonable percentage accuracy; and when the frequency was halved in order that a wide



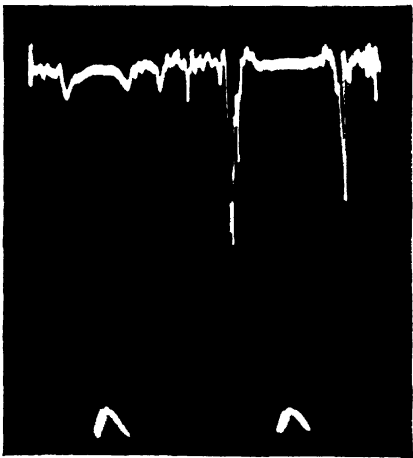
(i) Travel 0.55 cm, Max. $R=2400$



(ii) Travel 0.80 cm, Max. $R=3480$

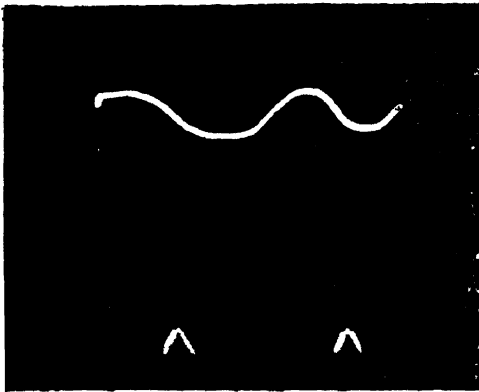


(iii) Travel 1.25 cm, Max. $R=5460$

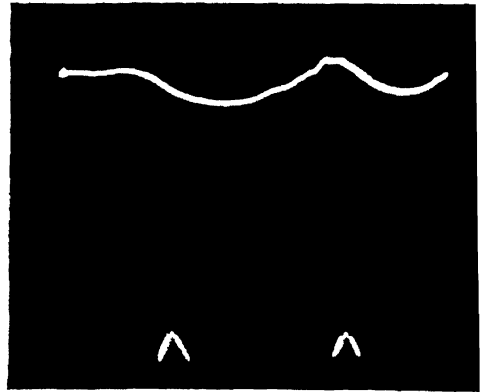


(iv) Travel 1.61 cm, Max. $R=6990$

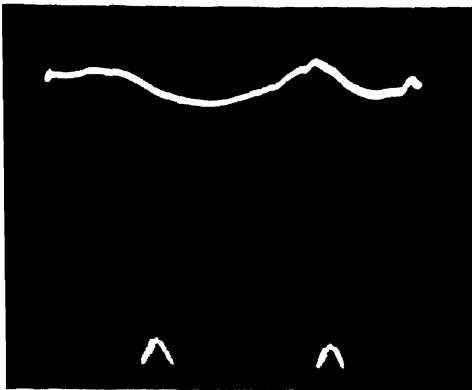
Figure 5 Breakdown of recurring oscillating flow



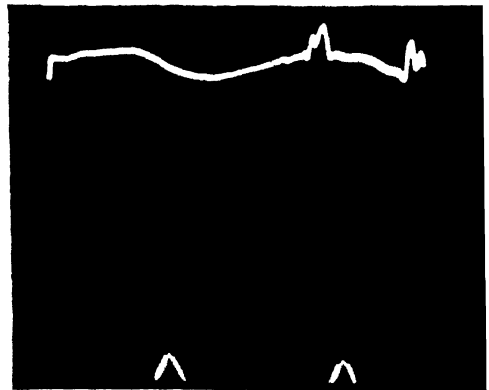
(i) Travel 4.39 cm, Max. $R=730$.



(ii) Travel 5.83 cm, Max. $R=980$



(iii) Travel 6.25 cm, Max. $R=1050$.



(iv) Travel 8.22 cm., Max. $R=1360$.

Figure 6. Breakdown of laminar oscillating flow.

range of conditions might be investigated, it approached so closely to the natural period that sufficiently small travels were unobtainable. Accordingly the aspirator bottles were replaced by vertical glass tubes, 19 in. long and only $1\frac{1}{8}$ in. in diameter, the lower portions of which were bent through an easy right angle. A carboy was substituted for the four capacities, and the resistance to the motion of the air was reduced by increasing the diameter of the connecting tubing from $\frac{1}{4}$ to $\frac{1}{2}$ in. The travel was measured with a horizontal telescope mounted on a cathetometer.

As expected, these modifications greatly reduced the damping, and tests were made at frequencies both above and below the natural period, but the difficulties mentioned above did not disappear. A test after three days at 20 c./min. showed, after a considerable wait, a smooth sinusoidal curve at $R=890$. At the next stage, a recurring trace with slight superposed ripples was obtained after allowing it a long time to settle, but the travel never became absolutely steady. Its mean was 2.51 cm. (corresponding to which R is 1050), but the excursions varied by about $\frac{1}{2}$ mm. at each end. These irregularities in the travel continued up to $R=1320$, accompanied by rippling of increasing violence, but at and beyond $R=1480$ no trouble was experienced in quickly obtaining both recurring traces and steady travels. Other observations made it clear that it was the Reynolds number, not the travel, which was the criterion of the onset of turbulence and of irregular travel. This conclusion is supported by the illustrations displayed in figure 6, which were taken with fluid seven days old at a speed of only $8\frac{1}{2}$ c./min. Although the sweep unfortunately gave trouble at this point, (i) shows that at $R=730$ the trace was smooth. The displacement of the trace being greater than in the three-day tests described above, at $R=800$ and 920 it was just possible to detect faint ripples, but they are too small to be seen on the necessarily small scale to which the photographs are reproduced here. At $R=980$, (ii), the rippling was more marked; at $R=1050$, (iii), however, the difficulty of obtaining steady travel again commenced, although the trace recurred. The irregularities were at their height at $R=1240$. Here the trace apparently settled down quickly, but then a spasm occurred and the trace slowly changed its form, not settling down to its original shape until 20 min. had elapsed. The recurring diagram, (iv), at $R=1360$ was taken after 15 min., the last 5 min. being steady, but at and beyond $R=1610$ no further trouble was encountered.

The results suggest a critical value of R , not very sharply marked, at about 800. The existence of a zone of irregular travel at slightly greater values of R may be attributed, not to the drive, but to the formation of small eddies, principally at the ends of the pipe. These would erratically increase the small viscous resistance to motion, and, when sufficiently large, would cause a noticeable variation in the travel. At greater amplitudes they would become more numerous and possess an almost constant mean value, and under these conditions a steady travel would again be obtained. It thus appears that the onset of turbulence can be detected, although in a less sensitive manner, simply by examining the uniformity of the travel.

§ 6 CONCLUSIONS

(i) Considerable difficulty was encountered in securing a satisfactory compromise between the conflicting requirements of the various parts of the apparatus. A wide pipe and a narrow beam of light were employed so that the lens effect of the former was small. Thus the emergent beam was sensibly parallel, and it was possible to obtain good extinction with the analyser. The disadvantage of the wide pipe was a very low critical velocity, occurring at such small shearing stresses that a slight anomaly in the viscosity of the working liquid, with its attendant distortion of the velocity distribution, was present. If, to avoid these effects, the working liquid had been a weaker solution, even more sensitive electrical apparatus would have been necessary.

(ii) With unidirectional laminar flow, the trace was a horizontal straight line. At a Reynolds number of 1970, brief disturbances appeared, which became more numerous and violent as the velocity was increased. Thus the indications of the apparatus were in fair accordance with the results obtained by other methods. Completely developed turbulence was shown to set in at a Reynolds number of about 2900.

(iii) With oscillating laminar flow, the trace was a smooth sinusoidal curve. At a certain Reynolds number (derived in an approximate manner) recurring ripples appeared, which suggested the formation of a steady system of vortices. Over a small range of slightly higher Reynolds numbers the motion of the liquid became slightly erratic owing, it is thought, to the formation of eddies which irregularly increased the viscous resistance. Under much more vigorous conditions of flow the recurring traces broke up, indicating complete turbulence, and this point was found to depend upon the travel of the liquid.

ACKNOWLEDGMENTS

The investigation described in the foregoing pages is the result of much cooperative effort. Mr. E. J. Bowen, F.R.S., suggested to me that the double-refraction effect was worth studying for the purpose of detecting turbulence, and he assisted in innumerable ways at every stage of the research. Mr. J. G. G. Hempson assembled the valve amplifier and the cathode-ray tube to his own design. Mr. V. Belfield carried out the whole of the photographic work, Dr. B. Lambert undertook the difficult task of jointing the 15-ft. glass tube, and Mr. W. M. Aitken assisted with the observations. To all these gentlemen I desire to express my grateful thanks. I am also greatly indebted to Messrs. Ricardo & Co., Ltd., for the loan of the reciprocating pump.

APPENDIX

Oscillatory horizontal displacements in the pipe

An expression is required giving the distribution of horizontal displacements over the cross-section of the pipe in terms of the amplitude of the surface movement in the bottles.

We adopt the notation used by Southwell and his collaborators (*loc. cit.*), and refer to their results by inserting the numbering of their equations at the left-hand margin of the page. Let p be the externally applied pressure, x the displacement in the bottles, v and y the horizontal velocity and displacement in the pipe at radius r . The uniform cross-section of the bottles is denoted by A , a is the radius of the pipe, μ and ρ are the viscosity and density of the liquid, $2h$ is the length of the pipe, and g is the acceleration due to gravity

We assume that

$$(4) \quad \left\{ \begin{array}{l} p = Pe^{mt}, \\ x = Xe^{mt}, \\ v = Ve^{mt}, \\ y = Ye^{mt}, \end{array} \right\} \quad \dots\dots(1)$$

where P , X , V and Y are complex, and V and Y are functions of r only. Then it was proved that

$$(8) \quad V = \frac{1}{k^2} \left(Q - \frac{P}{2h\mu} \right) \left[1 - \frac{J_0(kr)}{J_0(ka)} \right], \quad \dots\dots(ii)$$

$$(5) \quad \left\{ \begin{array}{l} \text{where} \\ \text{and} \end{array} \right. \quad \begin{array}{l} Q = \frac{2\pi g}{hA} \frac{k^2}{n^2} \int_0^a r V dr, \\ k^2 = -in\rho/\mu. \end{array}$$

$$(9) \quad \text{But} \quad X = \frac{h\mu}{g\rho} Q,$$

$$(10) \quad \text{and} \quad \left[\frac{g'}{hn^2} - \left\{ 1 - \frac{2}{ka} \frac{J_1(ka)}{J_0(ka)} \right\}^{-1} \right] X = \frac{g'}{g} \frac{P}{2\rho hn^2},$$

$$(7) \quad \text{where} \quad g' = \frac{\pi a^2}{A} g.$$

Hence, on integrating (ii), we obtain, after some reduction,

$$\frac{Y}{X} \frac{g'}{g} = \left\{ 1 - \frac{2}{ka} \frac{J_1(ka)}{J_0(ka)} \right\}^{-1} \left(1 - \frac{J_0(kr)}{J_0(ka)} \right) + E, \quad \dots\dots(iii)$$

where E is a function of r only. It is evidently zero when a steady state of oscillation has been attained. Now k^2 is a pure imaginary, hence we introduce a quantity z given by

$$(12) \quad z^4 = -k^4 a^4.$$

Then (iii) may be written

$$\frac{Y}{X} \frac{g'}{g} = \left\{ 1 - \frac{2}{iz\sqrt{z}} \frac{J_1(iz\sqrt{z})}{J_0(iz\sqrt{z})} \right\}^{-1} \left(1 - \frac{J_0(i\omega\sqrt{z})}{J_0(iz\sqrt{z})} \right). \quad \dots\dots(iv)$$

Here z and ω are real, and they are given by

$$z = a \left(\frac{n\rho}{\mu} \right)^{\frac{1}{4}}, \quad \omega = r \left(\frac{n\rho}{\mu} \right)^{\frac{1}{4}}.$$

The evaluation of the right-hand side of (iv) may be performed in two parts. The first term reduces to

$$(20) \left\{ \begin{array}{l} \text{where } C(z) = 1 + \frac{A(z) - 1}{\{A(z) - 1\}^2 + \{B(z)\}^2}, \quad D(z) = \frac{B(z)}{\{A(z) - 1\}^2 + \{B(z)\}^2}; \\ \text{and } A(z) = \frac{z W(z)}{2 V(z)}, \quad B(z) = \frac{z Z(z)}{2 V(z)} \end{array} \right. \dots (\bar{v})$$

The modulus of this vector is $[\{C(z)\}^2 + \{D(z)\}^2]^{\frac{1}{2}}$ and its phase angle θ is given by $\tan \theta = -D(z)/C(z)$. Numerical values of these expressions can be obtained with the aid of the tables given by Russell (1914). The second term in (iv), which gives the displacement distribution over the cross-section, may be expeditiously dealt with by writing it in the form

$$1 - \frac{M_0(\omega)}{M_0(z)} e^{i\{\theta_0(\omega) - \theta_0(z)\}},$$

and making use of McLachlan's tables (1934).

The diagrams of figure 4 can then be constructed. Taking OX to represent the direction of the vector X , we set off OP of length $[\{C(z)\}^2 + \{D(z)\}^2]^{\frac{1}{2}}$ at the phase angle θ . For various values of r/a , rays from P are drawn of length $[\{C(z)\}^2 + \{D(z)\}^2]^{\frac{1}{2}} M_0(\omega)/M_0(z)$, the angles between them and OP being $\{\theta_0(\omega) - \theta_0(z)\}$. Lastly, the curve joining the ends of the rays is drawn in.

REFERENCES

- ALCOCK, E. D. and SADRON, C. L., 1935. *Physics*, 6, 92.
 ANDRADE, E. N. DA C. and LEWIS, J. W., 1926. *Kolloidzshr.* 38, 260.
 CHRISTOPHERSON, D. G., GEMANT, A., HOGG, A. H. A. and SOUTHWELL, R. V., 1938. *Proc. Roy. Soc. A*, 168, 351.
 DEN HARTOG, J. P., 1940. *Mechanical Vibrations*, 2nd ed. p. 387 (New York and London. McGraw-Hill).
 EDSALL, J. T., 1942. *Advances in Colloid Science* (ed. Kraemer, E. O.), vol. 1, p. 269 (New York: Interscience Publishers).
 GOLDSTEIN, S. (ed.), 1938. *Modern Developments in Fluid Dynamics*, p. 319 (Oxford. Clarendon Press).
 HAUSER, E. A. and DEWEY, D. R., 1939. *Industr. Engng. Chem.* 31, 786.
 LAWRENCE, A. S. C., 1935. *Proc. Roy. Soc. A*, 148, 59.
 MARK, H., 1940. *High Polymers*, vol. 11, fig. 82 (New York: Interscience Publishers).
 McLACHLAN, N. W., 1934. *Bessel Functions for Engineers*, p. 182 (Oxford: Clarendon Press).
 RUSSELL, A., 1914. *The Theory of Alternating Currents*, 2nd ed. vol. 1, p. 233 (Cambridge: University Press).
 WELLER, R., MIDDLEHURST, D. J. and STEINER, D. J., 1942. *N.A.C.A. Technical Note*, no. 841.

EXACT ADDITION FORMULAE FOR THE AXIAL SPHERICAL ABERRATION AND CURVATURE OF FIELD OF AN OPTICAL SYSTEM OF CENTRED SPHERICAL SURFACES

By F. GILBERT BROWN,
Bexley

*Communicated by Prof L. C. Martin MS. received 4 February 1945 ;
in revised form 14 May 1945*

ABSTRACT. From a pair of axial magnification formulae, exact addition equations are derived in this paper for the axial spherical aberration and curvature of field for central and sagittal rays produced by any number of centred spherical surfaces whatever their individual separations.

For a ray of initial semi-aperture U_1 and final semi-aperture U_q' passing through media of refractive indices n_1, n_2, \dots, n_q , the condition for zero axial spherical aberration is

$$\sum_{p=q}^{p=1} n_p u_p c_p (\sin U_p + \sin I_p - \sin I_p' - \sin U_p') = 0,$$

where c_p is the distance of the p th intermediate paraxial object point from the centre of curvature of the p th surface

For zero Petzval curvature and an initial semi-angle of field θ , the condition is first given in the usual approximate Petzval form with the addition of a corrective term

$$\sum_{p=q}^{p=1} \frac{n_p' - n_p}{n_p' n_p r_p} \left\{ 1 + \tan^2 \frac{\theta_p}{2} - 2 \frac{n_p'}{n_p} m_p \left(\frac{x_p}{Y_p} - \tan \frac{1}{2} \theta_p \right)^2 \right\} = 0$$

A more simple alternative form, however, is $\sum_{p=q}^{p=1} \delta_p \tan \frac{1}{2} \theta_p = 0$, where δ_p is the deviation of the central paraxial ray at the p th surface.

An addition equation is also given for the sagittal astigmatism.

§ 1. INTRODUCTION

EXACT values for the aberrations of the image formed by compound optical systems of centred spherical refracting surfaces are usually obtained by trigonometrical ray-tracing, and the removal of the aberrations, although supplemented by well-known approximate formulae in many cases, thus depends in some degree on a trial-and-error process, in which values for the unknown radii of curvature are first assumed and then modified or adjusted until the selected aberrations are removed.

The object of this paper is to show that so far as the axial spherical aberration, the curvature of field for central rays (Petzval curvature), and to a lesser degree the sagittal astigmatism, are concerned, exact addition formulae can be derived, containing a minimum number of variables, which, when equated to zero, give the direct conditions for the absence of these aberrations.

It is hoped that the publication of these formulae may, in combination with

the well-known use of the sine condition for the pre-determination of coma, indicate useful means of estimating in advance the magnitude of the several aberrations in any given system, and at the same time encourage further development along these lines.

§ 2. THE AXIAL MAGNIFICATION FORMULAE

The formulae developed in this paper are based upon a pair of axial magnification formulae connecting (a) the successive intercepts Δl and $\Delta l'$ cut off respectively before and after refraction by a pair of paraxial rays, and (b) the corresponding intercepts ΔL and $\Delta L'$, when one of the pair is paraxial and the other of finite semi-aperture U .

For infinitesimal lengths dl and dl' it is well known that the relation for paraxial rays is

$$dl' = \frac{n'}{n} m^2 dl,$$

but for intercepts of finite length, this equation is an approximation only.

The exact equation of transfer for a pair of paraxial rays is

$$\Delta l' = \frac{n'}{n} m m_0 \Delta l \quad \text{or} \quad n' u' u_0' \Delta l' = n u u_0 \Delta l, \quad \dots\dots(1)$$

where m and m_0 are the linear magnifications at the respective extremities of the length $\Delta l'$, u and u_0 are the angles between the respective paraxial rays and the axis before refraction, and u' and u_0' are the corresponding angles after refraction. This equation may be deduced from the familiar paraxial formulae

$$ff' = -gg' = -g_0g_0' \quad \text{and} \quad f/g = m f'/g_0 = n' m_0/n,$$

remembering that $\Delta l' = g' - g_0'$, where the distances g , g_0 , g' , g_0' are measured from the principal foci f and f' respectively.

It is sometimes convenient to substitute m for m_0 , by putting $g + \Delta l$ for g_0 , so that

$$\frac{m}{m_0} = \frac{g + \Delta l}{g} = 1 + \frac{\Delta l}{g} = 1 + \frac{m \Delta l}{f},$$

whence

$$\frac{1}{m_0} = \frac{1}{m} + \frac{\Delta l}{f};$$

similarly

$$\frac{1}{m} = \frac{1}{m_0} - \frac{\Delta l}{f}. \quad \dots\dots(2)$$

Substituting the value of m_0 above in equation (1),

$$\Delta l' = \frac{n'}{n} m^2 \Delta l \left(1 - \frac{m_0 \Delta l}{f} \right), \quad \dots\dots(2b)$$

an equation which will be useful later on in the paper.

Next, let one of the rays be of finite aperture U cutting the optical axis in B and B' before and after refraction at distances C and C' from the centre of curvature of the refracting surface (see figure 1). Let b and b' denote the corresponding intersections of the paraxial ray at distances c and c' from the centre. Finally,

let b_0' represent the position of the paraxial image of B, and c_0' its distance from the centre of curvature.

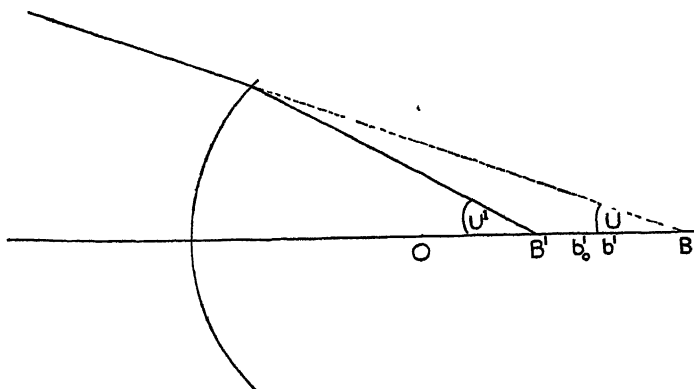


Figure 1

$$OB = C = c_0$$

$$OB' = C'$$

$$Ob = c$$

$$Ob' = c', \text{ paraxial image of } b$$

$$Ob_0' = c_0', \text{ paraxial image of } B$$

$$Bb = \Delta L = \Delta l$$

$$B'b' = \Delta L'$$

$$b_0'b' = \Delta l'$$

$$B'b' = A = \text{longitudinal spherical aberration}$$

Then, using $\Delta L'$ and ΔL in place of $\Delta l'$ and Δl in order to indicate that one of the image-forming rays is now of finite aperture, and denoting the magnifications at b' , b_0' and B' by m , m_0 and M respectively, we have, from equation (1),

$$\Delta L' = \frac{n'}{n} mm_0' \Delta L + A,$$

where A is the longitudinal spherical aberration of the image of the point B for semi-aperture U .

Substituting c_0'/C for m_0 in the preceding equation, we have

$$\begin{aligned} \Delta L' &= \frac{n'}{n} m \frac{c_0'}{C} \Delta L + A \\ &= \frac{n'}{n} m \frac{C' + A}{C} \Delta L + A \\ &= \frac{n'}{n} m M \Delta L + A \left(1 + \frac{n'}{n} m \frac{\Delta L}{C} \right), \text{ since } \frac{C'}{C} = m_0. \end{aligned}$$

Again, since $\frac{n'}{n} mm_0 \Delta L = \Delta l'$,

$$\frac{n'}{n} m \frac{\Delta L}{C} = \frac{c_0'}{c_0} \frac{\Delta l'}{c}.$$

Hence

$$\Delta L' = \frac{n'}{n} m M \Delta L + \frac{c'}{c_0} A = \frac{nu \sin U}{n'u' \sin U'} \Delta L + \frac{c'}{c_0} A. \quad \dots (3)$$

In a system of centred spherical refracting surfaces, the $\Delta L'$ of one surface becomes the ΔL of the next, and the U' of the one surface the U of the next. Equation (3) may therefore be extended to

$$n_q' u_q' \sin U_q' \Delta L_q' = n_1 u_1 \sin U_1 \Delta L_1 + \sum_{p=q}^{p=1} n_p' u_p' \sin U_p' A_p \frac{c_p'}{c_{0p}}. \quad \dots (4)$$

If in equation (+) a substitution can be made for the A_p in terms of the known elements of the p th surface, we have available an addition equation for spherical aberration. By a happy accident, as will be seen in the next section, we can not only substitute a simple expression for A_p , but also get rid of all the troublesome quantities of the type c'_{op} .

§ 3 SPHERICAL ABERRATION OF A RAY OF FINITE APERTURE U , PRODUCED BY A SINGLE SPHERICAL SURFACE, SEPARATED BY MEDIA OF REFRACTIVE INDICES n AND n'

In the present section, departing from accepted procedure, the radii of curvature of the refracting surfaces are treated as unknown quantities to be subsequently determined, and the formula for spherical aberration is based, in the first instance, on the semi-apertures U and U' before and after refractions. These are selected as known quantities and may be said to determine the *shape* of the system.

It only remains to fix a single linear constant (conveniently represented by paraxial c')-to determine the *scale*, after which all other linear quantities are uniquely determined, including the curvature of the surface and the spherical aberration

By definition we have

$$\frac{A}{C} = \frac{C' - c'}{C} = \frac{n \sin U}{n' \sin U'} - \frac{nu}{n'u'} = \frac{nu}{n'u'} \left(\frac{au' - \sin U'}{\sin U'} \right),$$

if $\sin U = au$.

If u be taken so that $a = 1$, the preceding equation becomes

$$A \sin U' = -c' (\sin U' - u'), \text{ since } nuc = n'u'c'.$$

Alternatively this may be written:

$$A \sin U' = c' (\sin U + \sin I - \sin I' - \sin U'), \quad \dots\dots(5)$$

for, as $a = 1$, $u' = u + i - i' = \sin U + \sin I - \sin I'$, where i and I represent the angles of incidence for the paraxial and marginal ray respectively, and i' and I' the corresponding values for the refracted ray.

The expression within the brackets will henceforth be designated as $\Delta(UU')$. Its value for any given ray of semi-aperture U and U' before and after refraction may be readily determined (by interpolation if necessary) from a set of specially prepared tables* in which the value of $\sin I - \sin I'$ is tabulated for standard values of the deviation $D = U' - U = I - I'$, and for the relative refractive index n'/n .

Alternatively $\sin I - \sin I'$ may be replaced by $i - i' = \delta$, so that $\Delta(UU') = \{(u + \delta) - \sin(U + D)\}$, the value of D , in terms of δ being given by the series

$$D = \delta - \frac{1}{2} \frac{1}{3} \frac{n'^3 - n^3}{(n' - n)^3} \delta^3 + \frac{1}{2} \frac{3}{4} \cdot \frac{1}{5} \frac{n'^5 - n^5}{(n' - n)^5} \delta^5 - \frac{1}{2} \frac{3}{4} \cdot \frac{5}{6} \cdot \frac{1}{7} \frac{n'^7 - n^7}{(n' - n)^7} \delta^7 \dots + \text{etc.} \quad \dots\dots(6)$$

For if we expand I in terms of $\sin I$, and I' in terms of $\sin I'$, we have

$$I = \sin I - \frac{1}{2} \cdot \frac{1}{3} \sin^3 I + \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{1}{5} \sin^5 I \dots \text{etc.},$$

and similarly for I' .

* The writer prepared some time ago a set of such tables, covering deviations between 0° and 12° and values of $\log n'/n$ between 18000 and 22000

Hence

$$D = I - I' = (i - i') - \frac{1}{2} \cdot \frac{1}{3} \left(\frac{n'^3 - n^3}{n'^3} \right) i^3 + \text{etc.}, \text{ since } \sin I' = \frac{n}{n'}, \sin I \text{ and } i = \sin I.$$

$$\text{Then equation (6) follows, since } \delta = i - i' = \frac{n' - n}{n'} i, \quad i^3 = \frac{n'^3}{(n' - n)^3} \delta^3; \\ \therefore - \frac{n'^5}{(n' - n)^5} \delta^5; \text{ and so on.}$$

The series in equation (6) is rapidly convergent for small values of δ , but for the rapid evaluation of $\Delta(UU')$ it is quicker to pick out from a table of logarithmic sines a pair of values for $\log \sin I$ and $\log \sin I'$, which differ by $\log n'/n$, while I and I' differ by the required deviation $U' - U$.

§ 4 AXIAL SPHERICAL ABERRATION OF THE MARGINAL RAY AFTER REFRACTION BY q -CENTRED SPHERICAL SURFACES

We can now substitute in equation (4) the value of A found in equation (5) above, noting that the c_0' in the former equation corresponds to the c' in equation (5) for each successive surface.

Hence

$$n_q' u_q' \sin U_q' A_q = n_1 u_1 \sin U_1 A_1 + \sum_{p=q}^{p=1} n_p' u_p' c_p' \Delta(U_p U_p'). \quad \dots (7)$$

In the absence of initial spherical aberration, the condition that the final ray U_q' shall be free from spherical aberration is

$$\sum_{p=q}^{p=1} n_p' u_p' c_p' \Delta(U_p U_p') = 0. \quad \dots (8)$$

Thus the general procedure in building up a system free from axial spherical aberration is first to define the path of the marginal ray through the system by specifying the U' of each successive surface (the U_p of the p th surface being identical with the U'_{p-1} of the preceding). These deviations having been chosen in accordance with the requirements of achromatism and the desired angular magnification of the system, factors are then selected representing the partial product $n_p' u_p' c_p'$ for each surface, in such a way as to make the algebraic sum on the left-hand side of equation (7) zero. This selection must of course be conditioned by other practical requirements of the system, such as the lens thicknesses, the separations of the individual lenses, and the intersection heights, as well as by the need to remove other aberrations.

§ 5 CONTROL OF COMA

Control of coma will normally be achieved by the well-known application of the sine condition. With an initial $u_1 = \sin U$, it will merely be necessary to see that final $u_q' = \sin U_q'$.

In the early stages of laying out the system, however, u_q' has not been determined, its exact value being dependent on the linear quantity c , and we only know the product uc . If, however, we write

$$k_p = n_p u_p c_p \quad \text{and} \quad K_p = \frac{\sum k_p \Delta(UU')_p}{\sin U_p'}$$

then the ratio of the magnification M_p of the marginal ray to the magnification m_p of the paraxial ray is

$$\frac{k_p(k_1 + K_1)(k_2 + K_2)(k_3 + K_3) \dots (k_{p-1} + K_{p-1})}{k_1(k_2 + K_1)(k_3 + K_2)(k_4 + K_3) \dots (k_p + K_{p-1})}, \dots (9)$$

which must be equal to unity if the final coma is to be zero.

$$\begin{aligned} \text{For } M_p/m_p &= \frac{\left(1 + \frac{A_1}{C_1'}\right)\left(1 + \frac{A_2}{C_2'}\right)\left(1 + \frac{A_3}{C_3'}\right) \dots \left(1 + \frac{A_p}{C_p'}\right)}{\left(1 + \frac{A_0}{C_1}\right)\left(1 + \frac{A_1}{C_2}\right)\left(1 + \frac{A_2}{C_3}\right) \dots \left(1 + \frac{A_{p-1}}{C_p}\right)} \\ &= \frac{\left(1 + \frac{K_1}{k_1}\right)\left(1 + \frac{K_2}{k_2}\right)\left(1 + \frac{K_3}{k_3}\right) \dots \left(1 + \frac{K_p}{k_p}\right)}{\left(1 + \frac{K_0}{k_1}\right)\left(1 + \frac{K_1}{k_2}\right)\left(1 + \frac{K_2}{k_3}\right) \dots \left(1 + \frac{K_{p-1}}{k_p}\right)}, \end{aligned}$$

which reduces to equation (9) when K_0 and K_p are each zero, which will always be the case if the initial and final spherical aberration be zero.

§ 6 SPECIAL CASE WHEN THE REFRACTING SURFACE IS PLANE OR OF EXCEPTIONALLY LARGE RADIUS OF CURVATURE

When r is very large or infinite, c is also very large or infinite, while $\sin I$ is equal or nearly equal to $-\sin U$ and $\sin I'$ to $-\sin U'$. Hence $\Delta(UU')$ tends to vanish identically. The following substitute formula for $nuc\Delta(UU')$ may then be used:—

Since $nuc = nrr$, and $\sin I = c \sin U/r$, $\sin I' = c' \sin U'/r$, we have, substituting in equation (5),

$$\begin{aligned} nuc(\sin U + \sin I - \sin I' - \sin U') &= nri\{\sin U(1 + c/r) - \sin U'(1 + c'/r)\} \\ &= n(L \sin U - L' \sin U'), \dots (10) \end{aligned}$$

the rs cancelling out.

This formula gives more accurate results for large radii and can be used for a plane surface. Otherwise it is less convenient as it involves two linear constants instead of one.

To conclude this section it may be observed that the advantages in the use of equation (6) will be found more apparent in the case of separated compound lenses than for thin systems of conventional type, where the possibilities of combination of glasses and surfaces have been thoroughly explored.

Even where trigonometrical methods are preferred, however, the quantity $nuc\Delta(UU')$ should still be formed for each surface parallel to the other work, since it is so readily calculated from the available elements. It shows at once how much each surface is contributing to the total spherical aberration, reveals the necessity for adjustments at an earlier stage, and shows to what extent the spherical aberration has accumulated beyond the power of subsequent surfaces to reverse.

§ 7. CURVATURE OF FIELD FOR CENTRAL RAYS

The formation of images by thin pencils passing through the centre of curvature of each refracting surface can seldom be realized in practice. The Petzval surface on which such theoretical images would be formed is, however, of considerable practical importance, since the distance of the real intersections of the

image-forming rays, before or behind this surface, is the measure of the degree of astigmatism introduced by the refracting system, and, as is well known, the formation of a plane anastigmatic image of a plane object requires as a first condition that the Petzval curvature shall also be zero.

The Petzval curvature for indefinitely thin pencils of central rays may be defined by the well-known formula

$$\frac{2x_q'}{n_q' Y_q'^2} = \frac{2x_1}{n_1 Y_1^2} + \sum_{p=1}^p \frac{n_p' - n_p}{n_p n_p' r_p},$$

where r_p is the radius of curvature of the p th refracting surface, x_1 and Y_1 are the rectangular coordinates of an extra-axial object point, referred to an origin B on the optical axis where it is cut by the object surface, and x_p' and Y_p' are the corresponding rectangular coordinates of the p th image of the point $x_1 Y_1$ referred to the p th paraxial image of B as origin.

This equation is an approximation only, and although for most conjugate positions of object and image and a fairly small angle of field it gives sufficiently accurate results, with the development of wide-angle systems and of less conventional types there would appear to be good grounds for replacing, wherever possible, the approximate Seidel conditions with exact formulae. Coma is already taken care of by the sine condition; an exact addition equation for axial spherical aberration has been supplied in a previous section; summations are now given for the Petzval and sagittal curvatures.

An exact form of the Petzval equation for a single surface follows from equation (2b), viz.,

$$\Delta l' = \frac{n'}{n} m^2 \Delta l \left(\frac{1 - m_0 \Delta l}{f} \right).$$

For, following the well known method for finding the Petzval equation, with B as origin, let x , Y (figure 2) be the coordinates of an extra-axial point in the

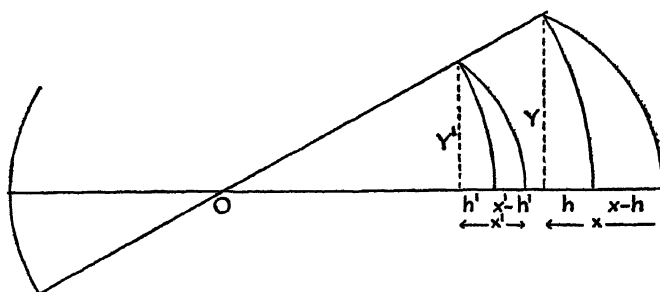


Figure 2.

object surface and x' , Y' the coordinates of the corresponding extra-axial image point formed by an indefinitely thin pencil of central rays. With centre at the centre of curvature of the refracting surface, describe arcs through the points xY and $x'Y'$. Let h and h' be the respective depths of curvature of these two arcs. We then have

$$\Delta l' = x' - h', \quad \Delta l = x - h, \quad Y'/Y = m.$$

Substituting in the above equation for $\Delta l'$ and Δl , we have

$$x' - h' = \frac{n'}{n} m^2 (x - h) \left(1 - m_0 \frac{x - h}{f} \right).$$

Dividing by $n'Y'^2 = n'm^2Y^2$ and transposing,

$$\frac{x'}{n'Y'^2} - \frac{x}{nY^2} = \frac{h}{Y} \left(\frac{1}{n'Y'} - \frac{1}{nY} \right) - m_0 \frac{(x-h)^2}{nY^2f},$$

since

$$h/Y = h'/Y'.$$

Remembering that $Y'/Y = C'/C$ and $C'/Y' = C/Y$, and that $\frac{Y^2+h^2}{Y^2} = \frac{2Ch}{Y^2}$,

$$\frac{x'}{n'Y'^2} - \frac{x}{nY^2} = -\frac{1}{2} \frac{n'-n}{nn'} \left(1 + \frac{h^2}{Y^2} - 2\frac{n'}{n} m_0 \frac{(x-h)^2}{Y^2} \right)$$

Again, if θ is the semi-angle of field, $\frac{h}{Y} = \tan \frac{1}{2}\theta$, so that, finally,

$$\frac{x'}{n'Y'^2} = \frac{x}{nY^2} - \frac{1}{2} \frac{n'-n}{nn'r} \left\{ 1 + \tan^2 \frac{1}{2}\theta - 2\frac{n'}{n} m_0 \left(\frac{x}{Y} - \tan \frac{1}{2}\theta \right)^2 \right\}. \quad \dots (11)$$

Since the $x'/n'Y'^2$ of one surface corresponds to the x/nY^2 of the next, the exact position of any image point formed by indefinitely thin central pencils in the Petzval surface after q refractions is given by the formula

$$\frac{x'_q}{n'_q Y'_q} = \frac{x_1}{n_1 Y_1} - \frac{1}{2} \sum_{p=2}^{p=q} \frac{n'_p - n_p}{n_p n'_p r_p} \left\{ 1 + \tan^2 \frac{1}{2}\theta_p - 2\frac{n'_p}{n_p} m_{0p} \left(\frac{x_p}{Y_p} - \tan \frac{1}{2}\theta_p \right)^2 \right\}. \quad \dots (12)$$

If the initial object surface is plane then $\frac{x_1}{n_1 Y_1}$ is zero.

The right-hand side of equation (11) then becomes

$$\frac{1}{2} \frac{n'-n}{nn'r} \left\{ 1 + \tan^2 \frac{1}{2}\theta \left(1 - 2\frac{n'}{n} m_0 \right) \right\}$$

This can vanish.

(a) For a plane refracting surface, i.e. $r = \infty$.

(b) When the object is at the centre of curvature of the refracting surface, i.e. when $m = n/n'$ and $\tan \frac{1}{2}\theta = 1$.

Combining these two conditions, it will be seen that the virtual image of a plane object formed by a plano-convex lens of focal length F at a distance $(n' - n)F/n'$ from the object has no Petzval curvature, although by the approximate formula the curvature is $1/nF$. This relative position of object and image is approximately realized at the field lens of a Huygenian eyepiece and in the front combination of some microscope objectives, thus assisting in the flattening of the field in such cases. Unfortunately, in a cemented plano-convex achromatic lens, the tendency of the crown is to neutralize this advantage, unless the refractive index of the glass of lower dispersion is at least as high as that of the flint.

Equation (11) consists of a principal term representing the usual Petzval sum with the addition of a corrective term for each surface. However, the exact condition that the Petzval sum may be zero can be expressed in a simpler form by dividing the axial magnification formula (1) by Y' instead of by Y'^2 and replacing Δl and $\Delta l'$ by $(x-h)$ and $(x'-h')$ as before.

Then for a single surface we have

$$u' \frac{x'}{Y'} = u \frac{x}{Y} + (u' - u) \frac{h}{Y} = u \frac{x}{Y} - \delta \tan \frac{1}{2}\theta, \quad \dots (13)$$

where δ is the paraxial deviation for a ray of semi-aperture u and u' before and after refraction.

For a succession of spherical surfaces the formula becomes

$$u_q' \frac{x_q'}{Y_q'} = u_1 \frac{x_1}{Y_1} - \sum_{p=q}^{p=1} \delta_p \tan \frac{1}{2} \theta_p, \quad \dots\dots(14)$$

so that if the object and final image surfaces are each to be plane,

$$\sum_{p=q}^{p=1} \delta_p \tan \frac{1}{2} \theta_p = 0. \quad \dots\dots(14a)$$

§8 ADDITION FORMULA FOR SAGITTAL ASTIGMATISM

By using equation (3) a condition similar to equation (13) above may be obtained for the curvature of the sagittal field.

Dividing by $n'Y' \sin V' = nY \sin V$, and again substituting $(X' - H')$ for $\Delta L'$ and $(X - H)$ for ΔL , and transposing, we have

$$\begin{aligned} u' \frac{X'}{Y'} &= u \frac{X}{Y} + (u' - u) \frac{H}{Y} + \frac{n'u'c' \Delta(VV')}{nY \sin V} \\ &= u \frac{X}{Y} - \delta \tan \frac{1}{2} \Psi + \frac{n'u'c' \Delta(VV')}{nY \sin V}, \quad \dots\dots(15) \end{aligned}$$

using capital X and capital H , V , V' and Ψ in place of x , h , u , u' and θ to distinguish the sagittal ray.

Then for a succession of q surfaces

$$u_q' \frac{X_q'}{Y_q'} = u_1 \frac{X_1}{Y_1} - \sum_{p=q}^{p=1} \delta \tan \frac{1}{2} \Psi + \sum_{p=q}^{p=1} \frac{n_p u_p' c_p' \Delta(VV')_p}{n_p Y_p \sin V_p}. \quad \dots\dots(16)$$

The final term in equation (16) has one serious disadvantage. It has already been stated that in the spherical-aberration formula (5), the scale of each element of the complete system is determined by a single linear constant c' . This constant again appears in equation (16), but there is no longer complete freedom of selection.

Although the scale of the system as a whole is still open, the relative scale as between the separate components has already been pre-determined by the selection (in conjunction with the paraxial deviation δ) of the angular subtense Ψ_p of the field presented to each surface, thus fixing the positions of the centres of curvature and leaving no remaining latitude in the radii after these have produced the prescribed deviations

It is therefore desirable to expand this term as follows:—

$$\begin{aligned} \frac{n'u'c' \Delta(VV')}{n'Y' \sin V'} &= \frac{nuc \Delta(VV')}{nY \sin V} = u \frac{\Delta(VV')}{\sin V} \left(\frac{C + X - H}{Y} \right) \\ &= u \frac{\Delta(VV')}{\sin V} \left(\operatorname{cosec} \Psi + \frac{X}{Y} - \tan \frac{1}{2} \Psi \right). \quad \dots\dots(17) \end{aligned}$$

For the preliminary work it will be sufficient to take $\frac{u}{\sin V} \Delta(VV') \operatorname{cosec} \Psi$ as the approximate value of the above term. The remaining corrections can be applied at a later stage.

Any combination of the second and third terms on the right-hand side of equation (16) will give a flat field for sagittal rays provided the total sum is zero; but to ensure that the sagittal image is formed in the Petzval surface, equation (14) must also be satisfied; or, a little less precisely, the two sums in equation (16) must be separately zero.

DYNAMIC MEASUREMENT OF YOUNG'S MODULUS FOR SHORT WIRES

By Y. L. YOUSEF,
Fouad I University, Cairo

MS. received 3 April 1945

ABSTRACT The method consists essentially in the measurement of the time for a quarter vibration of a short wire clamped at one end. The wire breaks a circuit as it starts its transverse vibration, and is made to strike a very light lever and open another circuit just as it completes one-quarter of its first vibration. The time between the two events is measured ballistically. From a knowledge of the time of vibration, the length, the diameter and the density of the wire, Young's modulus can be readily calculated. The method is particularly suitable for thin wires having a diameter of the order of 0.015 cm. and a length of about 2 cm. The results are consistent to a few per cent. Comparison of the results with the accepted values for five different wires shows a good agreement.

§1 INTRODUCTION

LORD RAYLEIGH (1926) correlated the frequency of transverse vibrations of a bar with its Young's modulus, and the correlation was used, probably for the first time, by Prosad (1929), who maintained the vibrations electrically, and determined the frequency chronographically by connecting to the free end an attachment for which an end-correction was applied. This correction amounted to more than 200% for lengths of the order of several cm. The mean results of Prosad differ by about 5% from the results he obtained by the flexural method, while his curves suggest that a divergence of about 15% from the average is possible.

In the present work, the frequency is calculated from the time of one-quarter of a vibration of the wire.

§2. EXPERIMENTAL DETAILS

A length of a few cm. of a thin straight wire is used. One end of the wire is rigidly clamped, and the other end can execute lateral vibrations. It is arranged that as the free end of the wire is released from a depressed position it breaks a circuit, and just as it reaches its normal position it opens another circuit. The time interval between the two events is measured by a condenser circuit described by Klopsteg (1920).

The principle will be understood by reference to figure 1. The two contacts K_1 and K_2 are opened in quick succession at the beginning and the end of the short interval to be measured. Then the battery is reversed, K_1 short-circuited by K_3 , and the two-way key K_5 put to the galvanometer side. If the setting of the resistances R and S has been right, the galvanometer throw will be zero, and the interval will be given by

$$t = CR \log_e R/S.$$

The contact K_1 . The wire X under test is clamped at C (figure 2) by means of two plane-edged aluminium jaws which can be tightened on the wire by nut and screw. The rigid clamp forms one terminal A of the contact. The clamped end of the wire is sharply defined by the plane edge of the clamp. The free end

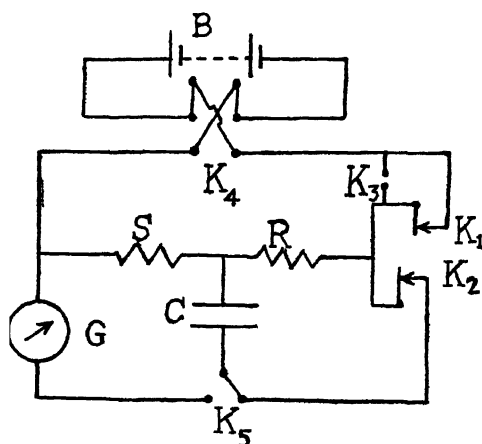


Figure 1. The principle of the method.

K of the wire can be depressed and kept in a constrained position by means of a stout nickel wire D having a small, flat, horizontal edge, with which it makes good electrical contact. D is connected to the second terminal B.

The contact at K is made at the extreme edges of X and D, so that when D is suddenly pulled outwards by means of an electromagnet or an attached thread,

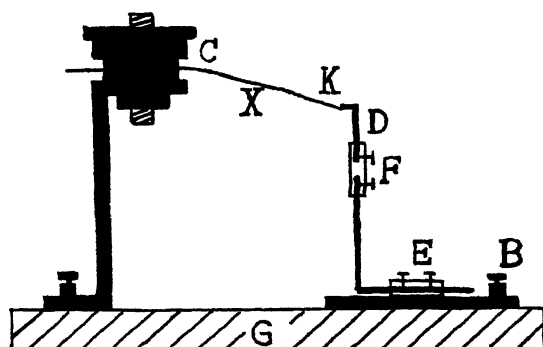


Figure 2 The contact K_1 .

the break of the battery-circuit of figure 1 is abrupt and the current drops immediately to zero.

The position of D can be adjusted both horizontally and vertically by means of the screws E and F, and thus measurements can be carried out for various lengths of X and for various amplitudes of vibration.

The arrangement is fitted on an ebonite base G, which itself is fixed centrally on a levelling table.

The contact K_2 . The contact K_2 has a very small inertia and at the same time has a reasonably low and fairly constant resistance (0.7 to 0.9 ohms).

JOL (figure 3) is a fine steel wire (diam. = 0.02 cm., length = 5 cm.) pivoted

at O where a pin passes through a small loop in the wire. The pin is soldered to a copper plate P connected to a terminal screw T_1 . The end L is bent at right angles to the length JL, and a short thin platinum wire (diam. = 0.0075 cm.) is soldered at N, where the lever rests on the plane top of the copper wire S, which is freshly but thinly amalgamated with mercury, and which forms the second

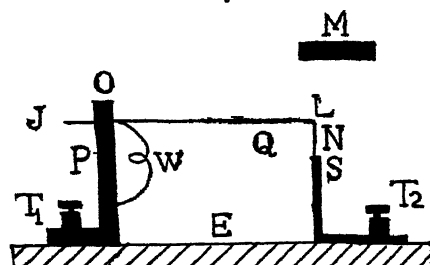


Figure 3. The contact K_2 .

lead. A very thin, loose wire W of copper (S.W.G. 40) is soldered between JL and the plate P to improve the contact between the lever and the terminal T_1 . It also helps, by the slight pressure it exerts, in keeping N and S in good contact, and thus the resistance between T_1 and T_2 is small. The whole arrangement is mounted on an ebonite base E, which is held by a fixed stand.

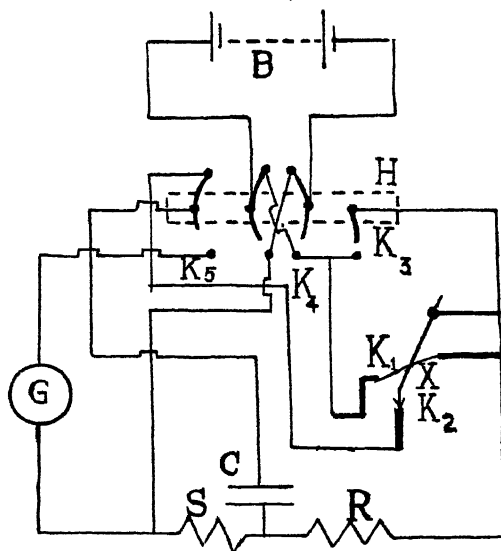


Figure 4. The experimental arrangement.

The vibrating wire X (figure 2) passes between E and JL, and is arranged to strike the lever normally at some point Q near the end L. As the lever receives the impulse, the contact between N and S is suddenly broken, and the point L is raised enough to come into the field of action of the small magnet M, to which it is attracted, and thus the contact remains open.

A force of a few centigrammes weight is sufficient to pull off the point N from S.

The actual arrangement. For K_3 , K_4 and K_5 of figure 1 we used mercury keys of the rocking type. They are constructed on one block of paraffin wax,

and are all operated by a common handle H (figure 4). The lengths of the movable links between the mercury pools are such that on switching over from the first position to the second, the short-circuiting key K_3 and the reversing key K_4 are closed just before the discharge K_5 .

R and S are two large resistance boxes ; C is a standard mica condenser of capacity $2\,0005 \times 10^{-6}$ F. ; G is a ballistic galvanometer having a sensitivity of 2 scale divisions per micro-coulomb, and a deflection of about 1/10 of a scale division can be detected ; the battery B has a constant e.m.f. of about 32 volts.

Adjustment of the test wire. The position of the unconstrained test wire, which is situated normally with its free end below the end of the lever, is carefully adjusted by levelling until it starts to raise the lever, and the contact K_2 is just broken. The break may be judged by viewing the tip N and its image on the bright amalgamated surface of S through a lens (figure 3). A white background assists in producing a distinct view. The right position may be located to within much less than corresponds to a 20° rotation by one of the screws of the levelling table supporting K_1 . This rotation is equivalent to a linear displacement of less than 1/200 of the smallest amplitude of vibration we used, which is about 3 mm.

On account of the very small inertia of the lever, the test wire is not appreciably bent as it starts to raise the lever. This is especially so for short lengths of the test wire, since the bending varies directly as the cube of the length.

When the "zero" position of the test wire is exactly located, it is depressed to the point D (figure 2) to close the contact K_1 . Care is always taken that the depression of the free end does not exceed 1/6 of the length used. For this depression, the initial amplitude of vibration may be regarded as small, and the wire will not change its zero position.

By virtue of the large velocity of vibration near the mean position, an error of $\pm 0.1\%$ of the amplitude arising from a slight error in locating the test wire will produce one of only $\frac{2}{\pi} \pm 0.1\%$ in the time of one-quarter vibration, or an error of $\frac{4}{\pi} \pm 0.1\%$ in Young's modulus

§ 3. THE METHOD OF MEASUREMENT

With K_2 closed (figure 4), observations are made by first approximately adjusting the position of the test wire, and then closing K_1 and taking trial values for R and S. On releasing the test wire, the contacts K_1 and K_2 will be opened in succession after a quarter vibration. The keys K_3 , K_4 and K_5 respectively are then immediately switched over by the handle H. If G shows a deflection, readjustment of the resistances R and S is made until the deflection is reduced practically to zero. It is generally sufficient to change one of the resistances only, and keep the other constant at a suitable value. The test wire is then exactly located, and the final adjustment of the resistance is made to give zero deflection exactly.

The values of R and S are now recorded, and a new balance is obtained with other values of the resistances, the zero position of the test wire being always checked during the final adjustments. An average time can thus be obtained.

If the contact K_2 has a small resistance r , then

$$t = C(R+r) \log_e \frac{R}{S}.$$

The frequency of transverse vibrations will be given by

$$\nu = \frac{1}{4t}.$$

Lord Rayleigh (loc. cit.) has shown that, for lateral vibrations of a bar of length l , density ρ and Young's modulus Y , the frequency is given by

$$\nu = \frac{\pi m}{2\pi l^2} \sqrt{(Y/\rho)},$$

where k is the radius of gyration of the section about an axis perpendicular to the plane of bending, and m is an abstract number whose value for the gravest mode of vibration of a clamped-free rod is 1.8751.

For a circular wire of radius a , $k = \frac{1}{2}a$, and hence

$$Y = \frac{12.77\rho}{a^2} \nu^2 l^4,$$

from which it will be seen that for a given wire $\nu^2 l^4$ is a constant quantity.

The length of the wire between the clamped and the free end is measured by callipers, and its average diameter is measured by a micrometer screw gauge. The density is determined from the mass and the dimensions of the sample; frequency determinations are made for various lengths of the wire, and the mean value of $\nu^2 l^4$ is obtained and used for calculating Y .

§ 4. RESULTS

Typical results are given in tables 1, 2 and 3, where the c.g.s. units are used throughout. The data in table 1 show the accuracy with which the

Table 1. Accuracy in measuring the time

R (ohms)	S (ohms)	t (sec.)	Average t	% deviation
1500	699	0.00229	0.00226	-1.4
1250	514	0.00222		+1.8
1000	322	0.00227		-0.4
750	168	0.00225		+0.4

Table 2. Effect of change of amplitude

A/l	R (ohms)	S (ohms)	t (sec.)	Mean t
0.222	2000	1010	0.002734	0.002738
	1500	602	0.002741	
0.109	2000	1013	0.002723	0.002715
	1500	609	0.002707	

Table 3. Results of the measurements

Material	ρ (gm./c.c.)	a (cm.)	l (cm.)	R (ohms)	S (ohms)	ν (sec. ⁻¹)	$\nu^{2/3} \cdot 10^{-5}$ (c.g.s. units)	$Y_0 \cdot 10^{-11}$ (c.g.s. units)	$Y_a \cdot 10^{-11}$ (c.g.s. units)
Brass	8.604	0.02283	5.29	2000	914	79.74	49.8	10.3	9.7 to 10.2
			4.52	1500	699	109.0	49.6		
			3.33	1000	535	199.5	48.9		
			2.68	500	219	302.0	47.1		
Steel	7.997	0.02793	5.32	1000	414	141.4	160.2	20.7	19.5 to 20.6
			4.68	1000	500	180.1	155.5		
			3.63	500	218	300.6	156.9		
			2.69	500	318	551.3	159.5		
Constantan	8.845	0.0273	5.535	1000	322	110.2	114.0	16.7	16.3
			5.385	1000	335	114.2	109.0		
			4.06	750	333	204.9	114.1		
			3.60	500	186	252.2	106.9		
Nickel	8.919	0.008675	2.605	750	273	164.8	12.50	19.6	20.2
			2.47	750	313	190.4	13.49		
			2.19	750	372	237.7	12.99		
			1.945	600	299	298.6	12.76		
Tungsten	19.8	0.00725	3.29	1500	564	85.9	8.48	39.9	39.0
			2.93	1000	306	105.5	8.19		
			2.74	1000	355	120.6	8.20		
			2.19	1000	519	190.4	8.32		
Nichrome	8.399	0.01802	5.13	1500	609	92.3	58.9	19.7	?
			3.41	1000	555	211.7	60.6		
			2.84	500	218	305.6	60.7		
			2.62	500	246	351.6	58.2		

adjustments can be made and the time measured. The observations were made on a constantan wire of length 5.535 cm. and radius 0.0273 cm.

The maximum departure from the average is less than 2%.

The figures in table 2 show the effect of change of amplitude on the time of a quarter vibration. The observations were made on a nichrome wire of length 5.13 cm and radius 0.01812 cm. The times for two widely different ratios of amplitude (A) to length (l) are given.

It will be seen that for the ratios of A/l given, the effect of change of amplitude does not exceed 1 %, which is of the same order of magnitude as the experimental error. However, the ratios of A/l we actually used in our determinations were always less than 0.16.

A summary of typical results at room temperature (18° c.) for six different wires is given in table 3, where Y_0 stands for Young's modulus by the author's method, and Y_a stands for the accepted value (Kaye and Laby, 1928; Childs, 1943)

The table shows how far the results of different observations on the same wire are concordant among themselves. It will be also noticed that, in general there is a good agreement between the observed and the accepted values of the elasticity.

§ 5 CONCLUSION

When measures are taken to secure good clamping of the test wire by a massive support, small inertia of the lever, high electrical conductivity at the contacts K_1 and K_2 , small area of contact with the tip of the lever, exact location of the test wire, good insulation of the various keys, and sudden break of the battery circuit by K_1 , the method appears to be very useful in measuring, with ordinary laboratory equipment, Young's modulus of elasticity, particularly for thin and short wires. No correction is required, and the wire need not be especially shaped or adapted. The average result will give a good measure of the true modulus of elasticity.

I wish to express my thanks to Dr. M. A. El-Sherbini for his valuable help and advice during the progress of this work.

REFERENCES

- CHILDS, W., 1943. *Physical Constants* (Methuen), p. 18.
 KAYE, G. and LABY, T., 1928 *Tables of Physical and Chemical Constants*, p. 29.
 KLOPSTEG, P., 1920 *Phys Rev* 15, 12.
 PROSD, K., 1929 *Phil Mag* 7, 548
 RAYLEIGH, Lord, 1926. *Theory of Sound*, 1, 273

A SYSTEM OF TRANSFER COEFFICIENTS FOR USE IN THE DESIGN OF LENS SYSTEMS: III. THE CONTRIBUTIONS TO THE IMAGE ABERRATIONS MADE BY THE INDIVIDUAL SURFACES OF A LENS SYSTEM

BY F. D. CRUICKSHANK,
University of Tasmania

MS received 26 February 1945

ABSTRACT A method is described which permits the computation of the contributions made by the individual surfaces of a lens system to the tangential aberrations of the final image for a pencil of finite aperture and obliquity. It employs the transfer coefficients previously described and forms part of a general development of the trigonometrical method of design. When applied to pencils of small aperture and small obliquity it gives an analysis of the surface contributions in exact agreement with known primary methods, and thus provides an extension from the analysis of the primary to the trigonometrically determined aberrations. In particular, expressions are derived for the contributions to the chromatic aberrations made by any component of the system which are simple functions of the dispersion of the component.

§ 1. INTRODUCTION

IN the course of the design of an optical system it is frequently of considerable value to know the contributions which each surface makes to the image aberrations. Methods of computing the surface contributions to the primary aberrations are well known, but, for aberrations other than these, no general analysis into surface contributions is available. Using the properties of the differential transfer coefficients developed in Part I of this paper, it is possible to solve this problem for the tangential aberrations of a system. We proceed to develop a trigonometrical method of computing the surface contributions characteristic of a lens system for a given aperture and position in the image field.

§ 2. THE CONTRIBUTIONS TO THE LONGITUDINAL AND TRANSVERSE CHROMATIC ABERRATIONS

In Part I an expression is derived for the longitudinal chromatic aberration of the system at a given zone in the form

$$Lch' = \Sigma \frac{\partial L_d'}{\partial N_h} (N_r - N_v)_h, \quad \dots\dots (67)$$

$$\text{where} \quad \frac{\partial L_d'}{\partial N_h} = \frac{\partial L_d'}{\partial n_2} \frac{1}{N_{h+1}} - \frac{\partial L_d'}{\partial n_1} \frac{N_{h-1}}{N_h^2}, \quad \dots\dots (68)$$

the summation being extended over all the components of the system. Each term in the summation in (67) represents the contribution made by a component to the longitudinal chromatic aberration of the final image. Thus for the general

component, h , the trigonometrically determined contribution to the longitudinal chromatic aberration, which we denote by $TLchC_h'$, is given by

$$TLchC_h' = \frac{\partial L_d'}{\partial N_h} (N_r - N_v)_h. \quad \dots\dots (69)$$

If the surface contributions are required rather than the component contributions, it is obvious that they may be easily obtained. Each term in the summation of equation (69) is derived from a pair of terms as seen in equation (68), the two members of each pair referring to single surfaces. Hence the longitudinal chromatic aberration may be expressed as a summation over the individual surfaces of the system if desired. For each air-glass surface there is a single term, while for a cemented surface there are two terms each of which involves the partial dispersion of one of the glasses forming the contact surface. The term, or the sum of the two terms, relating to each surface gives the contribution of that surface to the final value of Lch' . For the paraxial chromatic aberration we have similarly

$$lch' = \Sigma \frac{\partial l_d'}{\partial N_h} (N_r - N_v)_h, \quad \dots\dots (70)$$

whence
$$TlchC_h' = \frac{\partial l_d'}{\partial N_h} (N_r - N_v)_h. \quad \dots\dots (71)$$

In Part I an expression was also derived for the transverse chromatic aberration for a given obliquity in the form

$$Tch' = \Sigma \left[\frac{\partial H'_{prd}}{\partial N_h} \right]_{L'} (N_r - N_v)_h. \quad \dots\dots (72)$$

This expression is based on the trace of a principal ray of the desired obliquity using the refractive indices of the intermediate colour d . The subscript L' indicates that the plane in which the ordinates, H' , are measured is a fixed plane. It is convenient to use the paraxial image plane for this purpose. In a manner similar to that followed in the case of the longitudinal aberration, it may be shown that the trigonometrically determined contribution to the transverse chromatic aberration for the general component, h , is given by

$$TTchC_h' = \left[\frac{\partial H'_{prd}}{\partial N_h} \right]_{L'} (N_r - N_v)_h. \quad \dots\dots (73)$$

If the surface contributions are desired, the equation (72) may be rearranged as a summation over the surfaces in a manner similar to that already discussed in the case of the longitudinal aberration.

The contributions made by the components of the system to the final chromatic aberrations are frequently of great value in practical designing, being much more useful than the contributions made by the surface of the system. Their value lies in the fact that they are such simple functions of the dispersions, and it is often possible to improve the achromatism of the system by a change of glass at one or more components after inspection of the values of the contributions.

In the accompanying table, the results of the computation of the contributions made by the surfaces of a photographic objective to the paraxial longitudinal

chromatic aberration and to the transverse chromatic aberration at two obliquities are compared with the corresponding values for the contributions to the *primary* chromatic aberrations as calculated by the method of Conrady (1929).

Table

Surface	$TlchC'$	Conrady $lchC'$	$TTchC'$	Conrady $TchC'$	$TTchC'$	Conrady $TchC'$
				4°		10°
1	3 5406	3 5406	-0.0086	-0 0083	-0 0223	-0 0210
2	3.8759	3 8759	0 0778	0.0767	0.2138	0 1931
3	-6 6646	-6.6646	-0 1331	-0.1313	-0.3659	-0 3304
4	0 4251	0 4251	0 0723	0 0713	0.1949	0 1795
5	-2 0300	-2 0300	0.0101	0.0099	0 0271	0 0250
6	0 0052	0 0052	-0 0124	-0 0122	-0 0341	-0 0308
7	1.6591	1 6591	0.0034	0 0034	0 0089	0.0086

In the first place it will be seen that the contributions determined by the two methods lead to identical results for the longitudinal aberration. Secondly, the table shows that at an obliquity of 4° the values given by the two methods are in fairly close agreement, the small difference being due to aberrations of higher order than the primary, which are already present. As the obliquity increases, the two sets of values diverge further, as is to be expected. At very small obliquities the two sets of values would be in exact agreement. It will thus be seen that a perfectly general trigonometrical method of calculating the surface and component contributions for the chromatic aberrations of a system has been established. It is accurate, as may be checked by full ray-traces, and when applied to the paraxial region and to pencils of small obliquity it gives results in exact agreement with existing primary theory, thereby becoming a satisfactory extension of known methods.

§ 3. THE SURFACE CONTRIBUTIONS TO THE SPHERICAL ABERRATION

In the ray-trace of the axial pencil we have traced rays incident at certain zones of the lens system. In the image formed after refraction at surface i of the system, there is, for any such zone, a spherical aberration given by

$$LA_i' = l_i' - L_i',$$

which may be expressed in angular measure as

$$\begin{aligned} AA_i' &= LA_i' \sin U_i' / S_i' \\ &= LA_i' \sin U_i' \cos U_i' / (L_i' - X_i'). \end{aligned} \quad \dots\dots (74)$$

The angle AA_i' measures the departure of the ray from the ideal direction after refraction at the surface i . If the ray could be turned through an angle $dU_i' = -AA_i'$, the spherical aberration behind the surface i would be reduced to zero, and the effect at the final image would be the introduction of an amount of spherical aberration opposite to that introduced by the actual refractions over the first i surfaces of the system. In other words, the shift of the intersection point of the axial ray with the principal axis in the final image space due to this imagined

rotation of the ray through the angle $dU'_i = -AA'_i$, at surface i , will provide a measure of the sum of the spherical contributions of the first i surfaces of the system. We shall denote this quantity by the symbol $\sum^i TSC'$, the letters standing for the trigonometrically determined spherical contribution. Hence

$$\sum^i TSC' = \frac{\partial L'_k}{\partial U'_i} dU'_i = -\frac{\partial L'_k}{\partial U'_i} AA'_i, \quad \dots \dots (75)$$

and the individual contributions from the surfaces are given by

$$TSC'_i = \sum^i TSC' - \sum^{i-1} TSC'. \quad \dots \dots (76)$$

This gives a general trigonometrical method of analysing the surface contributions to the final spherical aberration. Little additional calculation is involved when the general methods outlined in these papers are employed, for the necessary transfer coefficients are already obtained in the main computation. The only limitation to the accuracy of the method is that differential transfer coefficients are used to calculate the effects in the final image-space of the rotations, dU' , of the ray. If the latter are large, the transfer coefficients will not predict their effects accurately. This difficulty is met to a very large extent by using appropriate second-order correction terms, as developed in the previous paper. If dU'_k is the change in the direction of the final emergent-ray due to the rotation of the ray through dU'_i behind surface i , then the shift of the intersection point along the principal axis is by equation (46).

$$\delta L'_k = C(U'_i) \operatorname{cosec} U'_k dU'_i (1 - \cot U'_k dU'_k), \quad \dots \dots (77)$$

and with due attention to the sign of the longitudinal aberration this leads to

$$\sum^i TSC' = C(U'_i) \operatorname{cosec} U'_k dU'_i (1 - \cot U'_k dU'_k). \quad \dots \dots (78)$$

This equation (78), incorporating the correction term, should take precedence over the earlier equation (75) and be used for the normal calculation of $\sum^i TSC'$. The separate surface contributions then follow from equation (76).

§4 THE SURFACE CONTRIBUTIONS TO THE DISTORTION

From the results of the trace of an oblique pencil through the lens system we can obtain the value of the distortion for this obliquity in the images formed after refraction at the successive surfaces of the system. In the usual linear measure this is

$$(\text{dist}')_i = (H'_{id} - H'_{prf})_i,$$

and in angular measure

$$\begin{aligned} AD'_i &= \text{dist}'_i \cos U'_{prn} / S'_{prfi} \\ &= \text{dist}'_i \cos^2 U'_{prn} / (l'_i - X'_{prn}). \end{aligned} \quad \dots \dots (79)$$

This angle, AD'_i , measures the departure of the principal ray from the ideal direction, as regards distortion, after refraction at surface i . If we could swing the principal ray through an angle $dU'_{prn} = -AD'_i$, we should restore the ray to its ideal direction and thereby remove from the final image the amount of distortion introduced by the refractions at the first i surfaces of the system. The effect of such an imagined rotation then is to introduce into the final image an amount

of distortion opposite to that contributed by the first i surfaces of the system, an amount which we denote by $-\sum TDC'$. Hence we have

$$\begin{aligned}-\sum TDC' &= d(H'_{id} - H'_{prf}) = -dH'_{prf} \\ &= -\frac{\partial H'_{prf}}{\partial U'_i} dU'_{prf} = \frac{\partial H'_{prf}}{\partial U'_i} AD'_i,\end{aligned}$$

that is,
$$\sum TDC' = -\frac{\partial H'_{prf}}{\partial U'_i} AD'_i, \quad \dots\dots(80)$$

and the contribution from the surface i is given by

$$TDC'_i = \sum TDC' - \sum^{i-1} TDC'. \quad \dots\dots(81)$$

Equation (80) is adequate provided the angles dU'_{prf} are small. Improved accuracy is obtained by using the second-order correction term; the difference in the amount of computation involved is so little that the corrected equation (82) given below should be used as a matter of course. Using equation (49) of the previous paper, and omitting the subscript pr , as all the symbols refer to the principal ray, we obtain finally

$$\sum TDC' = C(U'_i)_{prf} \sec U'_k dU'_i (1 + \tan U'_k dU'_k). \quad \dots\dots(82)$$

§ 5. THE SURFACE CONTRIBUTIONS TO THE TANGENTIAL COMA

We begin by computing the value of the tangential coma for a pencil of given obliquity in the images formed after refraction at the successive surfaces of the system. If we choose to regard the a and b rays of the pencil as aberrant, then it is the displacement of their intersection point, Ab , from the intersection point Pr which afflicts the system with coma. If after refraction at surface i we could swing the a and b rays until both passed through the point Pr , the tangential coma originally present in the image behind surface i would be removed. Such a swing of the rays would also leave the curvature of the tangential field unchanged. The effect of such an imagined rotation of the rays on the final image would be the introduction of an amount of tangential coma opposite to that contributed by the first i surfaces of the system, that is, an amount which we denote by $-\sum TCC'$. Thus we have a basis for the derivation of equations for computing the contribution of each surface of the system to the tangential coma of the final image.

The angle through which the a ray must be rotated is

$$\begin{aligned}dU'_{a_i} &= -\text{coma}'_{T_i} \cos U'_{a_i} / S'_{a_i} \\ &= -\text{coma}'_{T_i} \cos^2 U'_{a_i} / (L'_{ab} - X'_a)_i.\end{aligned} \quad \dots\dots(83)$$

The corresponding angle for the rotation of the b ray is

$$dU'_{b_i} = -\text{coma}'_{T_i} \cos^2 U'_{b_i} / (L'_{ab} - X'_b)_i. \quad \dots\dots(84)$$

In order to calculate the contributions at the final image we must follow the shifts of the intersection points Ab and Pr in the final image space due to these rotations. The shifts are

$$dL'_{ab} = \frac{\partial L'_{ab}}{\partial U'_{a_i}} dU'_{a_i} + \frac{\partial L'_{ab}}{\partial U'_{b_i}} dU'_{b_i}, \quad \dots\dots(85)$$

$$dH'_{ab} = \frac{\partial H'_{ab}}{\partial U'_{a1}} dU'_{a1} + \frac{\partial H'_{ab}}{\partial U'_{b1}} dU'_{b1}, \quad \dots\dots(86)$$

$$dH'_{pr} = -dL'_{ab} \tan U'_{prk}. \quad \dots\dots(87)$$

It then follows that

$$\sum TCC' = dH'_{ab} - dH'_{pr}, \quad \dots\dots(88)$$

and the contribution made by the surface i is given by

$$TCC'_i = \sum TCC' - \sum^{i-1} TCC'. \quad \dots\dots(89)$$

As in the previous case, it is generally advisable to use the second-order correction terms. It will be easily seen that the modifications necessary in equations (85) to (87) are that the first term on the right-hand side of (85) and (86) is multiplied by $[1 - \cot(U'_a - U'_b)_k dU'_{ak}]$, and the second term by the corresponding quantity $[1 - \cot(U'_a - U'_b)_k dU'_{bk}]$.

The essence of the foregoing method is the computation of the effect at the final image of a change made in the refracted rays behind surface i , the change being such that the coma present behind this surface is removed. Another change which would remove the coma in the image behind the surface i is a swing of the principal ray until it passes through the point Ab_i . It may seem more logical to make the changes in the directions of the rays a and b , because we regard coma as due to the aberrant behaviour of the outer rays of the pencil, but it is simpler to make the change of the principal ray for the purpose of calculating the coma contributions, as it involves less than half the amount of computation. The fact that the contributions computed either way are in close agreement confirms the validity of the general method. The angle through which the principal ray must be rotated to remove the coma behind surface i is

$$\begin{aligned} dU'_{pr1} &= \text{coma}'_{T1} \cos U'_{pr1} / S'_{pr1} \\ &= \text{coma}'_{T1} \cos^2 U'_{pr1} / (L'_{ab} - X'_{pr})_1. \end{aligned} \quad \dots\dots(90)$$

Following out the effect in the final image space of such a rotation we have

$$\sum TCC' = -\frac{\partial H'_{pr}}{\partial U'_{pr1}} \partial U'_{pr1}. \quad \dots\dots(91)$$

An additional advantage of the second method is the ease with which the correction term may be applied. As in the case of distortion, we obtain

$$\sum TCC' = -C(U'_i)_{pr} \sec U'_{prk} dU'_{pr1} (1 + \tan U'_k dU'_{prk}). \quad \dots\dots(92)$$

Finally, then, we use equations (90), (92) and (89) for the computation of the contributions to the tangential coma.

§6. THE SURFACE CONTRIBUTIONS TO THE CURVATURE OF THE TANGENTIAL FIELD

The curvature of the tangential field is measured usually by the distance of the intersection point Ab from a plane at right angles to the axis through the focus of some chosen axial ray. This ray is usually the extreme ray of an axial pencil having the same relative aperture as the oblique pencil. Thus

$$X_T' = L'_{ab} - L'_m.$$

As long as the point Ab lies in the plane through the intersection point of the ray m with the principal axis, the curvature of the tangential field would be zero according to the last equation, but the coma would be profoundly affected by the H' co-ordinate of the point. Considering the curvature as quite distinct from the coma, we could say that in a comatic system the ideal location of the point Ab as far as curvature is concerned is in the plane of the m focus at a point distant $(coma)_T'$ from the intersection point, Prm , of the principal ray with this plane. This analysis of coma and curvature may appear somewhat arbitrary, but so are most measures and analyses of the aberrations of a system. We propose to use the above ideas as a basis for determining the surface contributions to the tangential curvature.

From the results of the ray trace we calculate the tangential curvature of the image formed after refraction at successive surfaces of the system. Considering the rays a and b after refraction at surface i , a rotation of each ray so that their intersection point Ab_i now lies in the plane of the m focus at a point distant $(coma)'_{Ti}$ from the point Prm_i would remove the curvature originally present in the image behind the surface i . The effect of such imagined rotations would be to introduce at the final image an amount of curvature opposite to that introduced by the first i surfaces of the system, that is, an amount $-\sum TXC'$. This leads, as before, to a measure of the surface contributions.

The angle through which the a ray must be rotated is

$$\begin{aligned} dU'_{ai} &= X'_{Ti}(\tan U'_{ai} - \tan U'_{pri}) \cos U'_{ai}/S'_{ai} \\ &= X'_{Ti}(\tan U'_{ai} - \tan U'_{pri}) \cos^2 U'_{ai}/(L'_m - X'_a)_i, \quad \dots (93) \end{aligned}$$

with a corresponding expression for the rotation of the b ray given by

$$dU'_{bi} = X'_{Ti}(\tan U'_{pri} - \tan U'_{bi}) \cos^2 U'_{bi}/(L'_m - X'_b)_i. \quad \dots (94)$$

The effect of such changes in the directions of the rays on the curvature of the final image is given by

$$dX_T' = dL'_{ab} = \frac{\partial L'_{ab}}{\partial U'_{ai}} dU'_{ai} + \frac{\partial L'_{ab}}{\partial U'_{bi}} dU'_{bi}, \quad \dots (95)$$

and then

$$\sum TXC' = -dL'_{ab} \quad \dots (96)$$

and

$$TXC'_i = \sum TXC' - \sum^{i-1} TXC'. \quad \dots (97)$$

The correction term to be applied to equation (95) is exactly similar to that for equation (85), already noted.

REFERENCE

CONRADY, A. E., 1929. *Applied Optics and Optical Design* (London: Oxford University Press).

A SYSTEM OF TRANSFER COEFFICIENTS FOR USE IN THE DESIGN OF LENS SYSTEMS: IV. THE ESTIMATION OF THE TOLERANCES PERMISSIBLE IN THE PRODUCTION OF AN OPTICAL SYSTEM

By F. D. CRUICKSHANK,
University of Tasmania

MS. received 26 February 1945

ABSTRACT. A knowledge of the transfer coefficients for the aberrations permits the changes in the residual aberrations of the system to be calculated for any small departure from specified dimensions. Provided the changes in the residuals are small in comparison with the residuals themselves, the effect of the departure from the specified dimensions will be negligible. Hence the calculated values of the transfer coefficients provide a basis for the estimation of the limits within which each dimension of an optical system must be controlled in the workroom if systems having uniform characteristics are to be produced.

§1 INTRODUCTION

IN the production of an optical system, serious imperfections may be introduced by failure to achieve the curvatures and other dimensions specified in the design. On the other hand, the cost of production may be increased considerably by striving to attain the specified dimensions within limits which are unnecessarily fine. It is thus an important part of the designer's work to specify the limits within which each quantity must be controlled during production. It is proposed now to examine this problem in the light of the analysis developed in the preceding papers of this series, and it will be shown that a method of estimating tolerances of this kind may be obtained which links on logically and conveniently with the general method of final design which has been developed. It is to be emphasized that the tolerances under consideration are not those which are generally termed "optical tolerances", which have relation to the amounts of residual aberrations which may be permitted in a system which aims at a certain standard of definition in the image. The present considerations relate to the degree of control to be exercised in the optical shop in the production of a given system if a uniform product is to be produced.

The final design of a system always represents more or less of a compromise. It is characterized by certain residual aberrations which have been calculated during the final stages of the design, and which experience, or the performance of a carefully built prototype, shows to be compatible with satisfactory image definition. The whole success of known methods of design depends on the fact that the mathematical analysis of the aberrations gives a fairly reliable guide to the actual physical aberrations of the system when constructed. In the course

of production, any variation from the specified dimensions will result in a variation in the actual residual aberrations, but production will be satisfactory provided these variations are small compared with the residual aberrations themselves. Hence it is reasonable to base a system of estimating tolerances on the calculation of the effect on the residuals of a departure from the specified value of each quantity in the system. The reliability of these tolerances will be of the same order as the reliability of the design methods as a whole. The system of transfer coefficients developed in Part I provides the machinery for the method, little additional computation being required.

§ 2. THE TOLERANCES FOR THE CURVATURE OF EACH SURFACE OF THE SYSTEM

In Part I it is shown that it is possible to calculate a transfer coefficient which measures the rate of change of any aberration with the curvature of any surface of the system. These coefficients are calculated normally for the purpose of the final differential correction of the system. They are now available for the further purpose of estimating the tolerances permissible during production.

If we use the symbol A' to denote any of the tangential aberrations of the system, then the calculated values of $\partial A'/\partial c$ reveal the effect of small changes of curvature on this aberration. After a close scrutiny of these quantities, tolerances may be set by assigning a permissible curvature variation at each surface such that the sum over all the surfaces of the effects on the residual aberrations due to the occurrence of such variations shall not exceed some specified fractions of the values of the residual aberrations. If a curvature tolerance, $\pm dc$, is selected at each surface in this way, the corresponding tolerance in the radius of curvature is $\pm r_i^2 dc$. It is often convenient to express the tolerance in terms of the number of fringes across a surface of a certain diameter. It is easily shown that the difference in curvature, dc , between two spherical surfaces of curvatures c and $c + dc$, which show x fringes across a surface of diameter $2a$ when placed together, is given very closely by

$$dc = \frac{x\lambda}{a^2} (1 - 0.5 a^2 c^2). \quad \dots\dots (98)$$

Writing $\partial A'/\partial c_f$ for the change in the aberration A' per fringe change in curvature, we have for any surface

$$\frac{\partial A'}{\partial c_f} = \frac{\partial A'}{\partial c} \frac{\lambda}{a^2} (1 - 0.5 a^2 c^2). \quad \dots\dots (99)$$

Equation (99) provides the necessary basis for the estimation of the curvature-tolerances.

§ 3. THE TOLERANCE IN THE AXIAL THICKNESS OF A COMPONENT OR AIR SPACE

The thickness of a lens is a quantity which is much more difficult to control in production than the curvatures of its surfaces, so that the question of tolerances becomes very important. For any component the transfer coefficient, $\partial A'/\partial d$, for its rear surface specifies the change of the aberration per unit change of axial thickness. In accordance with the signs we have used, a negative d -change

increases the thickness of the component. Formally, then, we introduce an axial thickness coefficient, $\partial A'/\partial t$, defined by

$$\frac{\partial A'}{\partial t_h} = - \left(\frac{\partial A'}{\partial d_2} \right)_h, \quad \dots\dots(100)$$

the subscript h referring to the general component and the subscript 2 denoting the second, or rear, surface of that component. Thus, without further computation, a set of coefficients is available for the estimation of the tolerances for the thicknesses of the components and air spaces. We make use of them by assigning to each component and air space a permissible thickness variation such that the sum over all the components of the effects on the residual aberrations due to such variations shall not exceed some small fraction of the measures of the residual aberrations.

§ 4. THE TOLERANCES FOR THE REFRACTIVE INDEX AND DISPERSION OF THE GLASSES

It was shown in Part I that for any monochromatic aberration, A' , there may be calculated a differential coefficient, $\partial A'/\partial N_h$, which measures the change of this aberration in the image formed by the system per unit change of refractive index of the component, h , of the system. These coefficients provide full information as to the limits within which the refractive index of each component must be controlled for satisfactory production of the system, and permit a system of tolerances to be established. As before, we assign a permissible variation of the refractive index to each component such that the sum over all the components of the effects on the residual aberrations due to the permitted variations is less than a prescribed small fraction of these residuals. As regards the variation of the dispersions of the glasses, the coefficients, $\partial Lch'/\partial P_h$ and $\partial Tch'/\partial P_h$, developed in Part I furnish the basis for tolerances, but generally speaking, the variation of the dispersion of a glass between successive melts is negligible, so that these tolerances are seldom required. Frequently a more serious effect of the variation of refractive index is the resulting change in focal length of the system. This is dealt with in the next section.

§ 5. THE TOLERANCES FOR THE CONTROL OF THE FOCAL LENGTH OF THE SYSTEM

It is often important to ensure that the focal length of a particular system is controlled within fine limits during production. For this we must know the effect of the variation of curvature, thickness, and refractive index on the focal length. From the results of the paraxial ray trace

$$f' = y/u_k'$$

from a ray incident parallel to the axis at a height y on the first surface. Hence

$$\begin{aligned} \frac{\partial f'}{\partial c_i} &= - (y/u_k'^2) \frac{\partial u_k'}{\partial c_i} \\ &= - (f'/u_k') \frac{\partial u_k'}{\partial c_i}. \end{aligned} \quad \dots\dots(101)$$

Corresponding expressions hold for $\partial f'/\partial n_k$ and $\partial f'/\partial d_k$. For the effect of the variation in the thickness of a component we have

$$\frac{\partial f'}{\partial t_k} = (f'/u_k') \left(\frac{\partial u_k'}{\partial d_k} \right)_k, \quad \dots\dots(102)$$

and the effect of refractive-index variation is given by

$$\frac{\partial f'}{\partial N_k} = -\frac{\partial f'}{\partial n_1} \frac{N_{k-1}}{N_k^2} + \frac{\partial f'}{\partial n_2} \frac{1}{N_{k+1}}. \quad \dots\dots(103)$$

The transfer coefficients in equations (101) to (103) are quickly computed as the values of $\partial u_k'/\partial c_i$, etc., have been calculated in the general computation for the paraxial ray. Thus a series of tolerances may be set quite simply for the control of the focal length of the system.

§6 THE USE OF COMPONENTS OUTSIDE THE SPECIFIED TOLERANCES

When a large number of optical systems of a certain type are being produced, a certain percentage of rejected components with dimensions outside the specified tolerances is inevitable. The intelligent use of the information provided by the transfer coefficients calculated during the design enables some at least of these rejects to be sorted into sets of optics which will give a satisfactory performance. Components rejected on account of non-spherical figure are excluded from our considerations

The rejected components are classified according to the type and amount of their departure from specified dimensions. For each aberration a limit is set for the additional amount of this aberration which can be admitted on account of departure from exact specifications. This will be some small fraction of the calculated residual aberration. A table is prepared which sets out the aberration changes for departure from the specified dimensions in terms of the units used in the workroom. The trained worker then derives from this table the amount of each aberration introduced into the system if a certain reject component, A, is employed. A study of the transfer coefficients in the table will now reveal whether some departure from dimensions of another component, B, will introduce amounts of the various aberrations which will compensate those introduced by A, and thus bring their totals within the limits prescribed. If this is possible, a reject of type A is paired with B, etc., and in this way use may be made of the rejected components. This kind of attempt to salvage components is probably made in every optical shop by trial-and-error methods or by bench tests. The use of a table of coefficients as suggested organizes such attempts intelligently, giving a reliable guide to what is possible in this respect.

A SYSTEM OF TRANSFER COEFFICIENTS FOR USE IN THE DESIGN OF LENS SYSTEMS: V. TRANSFER COEFFICIENTS FOR THE ASTIGMATISM AT SMALL APERTURE AND FINITE OBLIQUITY

By F. D. CRUICKSHANK,
University of Tasmania

MS. received 26 February 1945

ABSTRACT Transfer coefficients are developed which specify the changes of the positions of the tangential and sagittal foci in narrow pencils of any obliquity due to small changes in curvature, refractive index, and axial separation which may be made within the system

§1 INTRODUCTION

IN the preceding papers of this series, transfer coefficients have been developed for the tangential aberrations only of a lens system. To complete the development it is necessary to establish a means of handling the astigmatism in a similar manner. The full solution of this problem, however, requires the extension of the foregoing methods to the general skew trace. Not having had opportunity to give much attention to this problem as yet, the writer has used a simple analysis of the astigmatism at small aperture in pencils of any obliquity, checking the astigmatism at full aperture from time to time by the usual skew trace. This constitutes quite a useful designing tool, as the changes in the astigmatism at small aperture due to changes within the system can be estimated for the differential correction process. The transfer coefficients for the shifts of the foci of narrow fans of tangential and sagittal rays are deduced, and this provides a basis for estimating the astigmatism changes produced by any alteration made within the system.

§2 THE TRANSFER COEFFICIENTS FOR THE SHIFT OF THE TANGENTIAL FOCUS

In addition to the principal ray of the oblique pencil of given obliquity, we trace another ray of the same obliquity close to the principal ray and calculate for it the single surface coefficients and the transfer coefficients in the normal way. We denote this ray by the subscript *cp*, the letters standing for *close principal* ray. From the trace of the principal ray we have, as in Part I, equation (45 a),

$$t_k' = \frac{\partial p_k'}{\partial p_1} \bigg/ \frac{\partial U_k'}{\partial p_1}$$

The two rays pr and cp will emerge from the last surface of the system at different points, and it is obvious that we may specify the separation of the emergence points by

$$\delta p'_{prk} = t'_k (U'_{cp} - U'_{pr})_k = t'_k \delta U'_k. \quad \dots\dots (104)$$

If a change is now made at some surface within the system, the two traced rays will emerge from points with a different separation and will intersect in the final image-space under a new angle. Thus if we can calculate the changes in $\delta p'_{prk}$ and $\delta U'_k$ resulting from the alteration made within the system, we can calculate the new value of t'_k specifying the position of the tangential focus.

Let us suppose that a small curvature change is to be made at surface i within the system. Differentiating equation (104),

$$\frac{\partial}{\partial c_i} (\delta p'_{prk}) = t'_k \frac{\partial}{\partial c_i} (U'_{cp} - U'_{pr})_k + \frac{\partial t'_k}{\partial c_i} (U'_{cp} - U'_{pr})_k.$$

For the term on the left-hand side of this equation we have

$$\frac{\partial}{\partial c_i} (\delta p'_{prk}) = \frac{\partial p'_{cpk}}{\partial c_i} - \frac{\partial p'_{prk}}{\partial c_i},$$

and combining these two equations we obtain

$$\frac{\partial t'_k}{\partial c_i} = \frac{1}{\delta U'_k} \left[\left(\frac{\partial p'_{cpk}}{\partial c_i} - \frac{\partial p'_{prk}}{\partial c_i} \right) - t'_k \left(\frac{\partial U'_{cpk}}{\partial c_i} - \frac{\partial U'_{prk}}{\partial c_i} \right) \right]. \quad \dots\dots (105)$$

The corresponding expressions for thickness changes and refractive-index changes may be written down, giving

$$\frac{\partial t'_k}{\partial d_i} = \frac{1}{\delta U'_k} \left[\left(\frac{\partial p'_{cpk}}{\partial d_i} - \frac{\partial p'_{prk}}{\partial d_i} \right) - t'_k \left(\frac{\partial U'_{cpk}}{\partial d_i} - \frac{\partial U'_{prk}}{\partial d_i} \right) \right], \quad \dots\dots (106)$$

$$\frac{\partial t'_k}{\partial n_i} = \frac{1}{\delta U'_k} \left[\left(\frac{\partial p'_{cpk}}{\partial n_i} - \frac{\partial p'_{prk}}{\partial n_i} \right) - t'_k \left(\frac{\partial U'_{cpk}}{\partial n_i} - \frac{\partial U'_{prk}}{\partial n_i} \right) \right]. \quad \dots\dots (107)$$

It will thus be seen that the transfer coefficients for the t -focus require little calculation beyond the work involved in the trace and coefficients of the ray cp . If changes are made during the process of differential correction which are too large to be treated as differentials, an improvement in accuracy may be obtained by calculating the new values of $\delta p'_k$ and $\delta U'_k$ resulting from the change and deducing the new value of t'_k directly from them. Thus, for a curvature change δc_i , we have

$$r (\delta p'_k) = \delta p'_k + \left(\frac{\partial p'_{cpk}}{\partial c_i} - \frac{\partial p'_{prk}}{\partial c_i} \right) \delta c_i, \quad \dots\dots (108)$$

$$\text{new } (\delta U'_k) = \delta U'_k + \left(\frac{\partial U'_{cpk}}{\partial c_i} - \frac{\partial U'_{prk}}{\partial c_i} \right) \delta c_i, \quad \dots\dots (109)$$

and the change in the t -focus is given by

$$\delta t'_k = \frac{(\delta p'_k)_{\text{new}}}{(\delta U'_k)_{\text{new}}} - t'_k. \quad \dots\dots (110)$$

§ 3 THE EFFECT OF CHANGES AT A SINGLE SPHERICAL SURFACE ON THE FOCUS OF A NARROW FAN OF SAGITTAL RAYS

The position of the sagittal focus behind the successive surfaces of a system is easily calculated by the well-known s -trace which is associated with the ordinary trace of the principal ray of the oblique pencil under consideration. We propose to use the normal relations of the s -trace to develop the transfer coefficients for the sagittal focus. The position of the sagittal focus behind any surface of the system is given by

$$\frac{N'}{s'} = \frac{N' \cos I' - N \cos I}{r} + \frac{N}{s}. \quad \dots\dots(111)$$

If we now write

$$\sigma = 1/s, \quad \sigma' = 1/s', \quad \dots\dots(112)$$

and

$$G = (N' \cos I' - N \cos I)/N', \quad \dots\dots(113)$$

equation (111) becomes

$$\sigma' = G\sigma + n\sigma. \quad \dots\dots(114)$$

This procedure clears the basic equation of reciprocals and makes it easier to handle in the subsequent differentiations. Let us suppose now that a small curvature change, dc_i , is to be made at the surface i of the system. The effect on the s -focus behind the surface will be determined by the derivative

$$\frac{\partial \sigma'_i}{\partial c_i} = G_i + c_i \frac{\partial G_i}{\partial c_i} + n_i \frac{\partial \sigma_i}{\partial c_i}.$$

The change σ_i due to a small change of curvature is of second order only, so that we may neglect the last term on the right-hand side of the above equation, and hence

$$\frac{\partial \sigma'_i}{\partial c_i} = G_i + c_i \frac{\partial G_i}{\partial c_i}. \quad \dots\dots(115)$$

To be able to compute the derivative $\partial \sigma'_i / \partial c_i$ we require an expression for $\partial G_i / \partial c_i$. This may be obtained by differentiation of equation (113) and the refraction equation $N \sin I = N' \sin I'$. Thus,

$$\begin{aligned} \frac{\partial G}{\partial I} &= -\sin I' \frac{\partial I'}{\partial I} + n \sin I \\ &= -\sin I' n \frac{\cos I}{\cos I'} + n \sin I \\ &= \sin I' \left(1 - n \frac{\cos I}{\cos I'} \right) \\ &= \sin I' \left(1 - \frac{\partial U'}{\partial U} \right). \end{aligned} \quad \dots\dots(116)$$

Differentiation of the computing equation

$$(L-r) \sin U = r \sin I$$

gives

$$\frac{\partial I}{\partial c} = L \sin U / \cos I,$$

and hence we have

$$\begin{aligned} \frac{\partial G}{\partial c} &= \frac{\partial G}{\partial I} \frac{\partial I}{\partial c} \\ &= \sin I' \left(-1 \frac{\partial U'}{\partial U} \right) \frac{L \sin U}{\cos I} \\ &= \sin I' \frac{\partial U'}{\partial c}. \end{aligned} \quad \dots\dots(117)$$

The computation of $\partial G/\partial c$ at any surface thus involves only one operation beyond the calculation of the ordinary single surface coefficients, which will have been computed already for the principal ray. Two further operations complete the calculation of $\partial \sigma'/\partial c$.

To determine the effect of a small change in refractive index at any surface i of the system, differentiation of equation (114) gives

$$\frac{\partial \sigma'_i}{\partial n_i} = \sigma_i + c_i \frac{\partial G}{\partial n_i}, \quad \dots\dots(118)$$

in which, by differentiation of (113), we have

$$\begin{aligned} \frac{\partial G}{\partial n} &= -\sin I' \frac{\partial I'}{\partial n} - \cos I \\ &= -\sin I' \frac{\sin I'}{\cos I'} - \cos I \\ &= \sin I' \frac{\partial U'}{\partial n} - \cos I. \end{aligned} \quad \dots\dots(119)$$

In this case only four operations are required to calculate $\partial \sigma'/\partial n$ for any surface after the single surface coefficients.

Finally, if a small change is made in the axial separation of two consecutive surfaces of the system, differentiation of equation (114) gives

$$\frac{\partial \sigma'_i}{\partial d_i} = c_i \frac{\partial G_i}{\partial d} + n_i \frac{\partial \sigma_i}{\partial d_i}. \quad \dots\dots(120)$$

Expressions for the two derivatives on the right-hand side of this equation may be deduced quite easily. Thus, differentiating equation (113),

$$\begin{aligned} \frac{\partial G}{\partial d} &= -\sin I' \frac{\partial I'}{\partial d} + n \sin I \frac{\partial I}{\partial d} \\ &= \sin I' \left(\frac{\partial I}{\partial d} - \frac{\partial I'}{\partial d} \right) \\ &= \sin I' \left(\frac{\partial I}{\partial d} - n \frac{\cos I}{\cos I'} \frac{\partial I}{\partial d} \right) \\ &= \sin I' \left(1 - \frac{\partial U'}{\partial U} \right) \frac{\partial I}{\partial p} \frac{\partial p}{\partial d} \\ &= \sin I' \sin U \frac{\partial U'}{\partial p}. \end{aligned} \quad \dots\dots(121)$$

To obtain an expression for the second derivative we differentiate the transfer formula of the s -trace, which is

$$s_i = s'_{i-1} - D'_{i-1}, \quad \dots\dots(122)$$

and thus

$$\begin{aligned} \frac{\partial \sigma_i}{\partial d_i} &= -\frac{1}{s_i^2} \frac{\partial s_i}{\partial d_i} \\ &= \sigma_i^2 \frac{\partial D'_{i-1}}{\partial d_i} \\ &= -\sigma_i^2 \sec U_i \left[1 - \frac{\sin U_i \sin (U+I)_i}{\cos I_i} \right]. \end{aligned} \quad \dots\dots(123)$$

The last step follows simply from the well known relation between D'_{i-1} and d_i .

Often the bracketed expression may be replaced by unity with a good degree of approximation, and is accurately unity for a plane surface. Equations (120), (121) and (123) permit the calculation of $\partial\sigma'/\partial d$ at any surface of the system.

§4 THE TRANSFER COEFFICIENTS FOR THE SAGITTAL FOCUS OF A NARROW PENCIL

It now remains to develop a method of computing the transfer coefficients for σ' -changes so that the effect of a change within the system on the final value of σ'_k may be determined. Since G is a function of the angles of incidence and refraction, its value at any surface will change with change of the point of incidence. It will be found, however, that in general the change of G along the path of the pencil due to small alterations within the system has a very small effect on the transfer coefficients, and, on account of the simplification obtained by neglecting this factor, it is proposed to develop a set of transfer coefficients which leave it out of consideration altogether.

Suppose that by some means a change $d\sigma'_i$ is produced in the narrow pencil refracted at surface i . To determine the effect of this change at the following surface of the system we require an expression for the derivative $\partial\sigma_{i+1}/\partial\sigma'_i$. We may write the transfer formula in (122) in the form

$$\sigma_{i+1} = \sigma'_i(1 - D'_i \sigma'_i)^{-1}, \quad \dots\dots(124)$$

and hence

$$\frac{\partial\sigma_{i+1}}{\partial\sigma'_i} = (1 - D'_i \sigma'_i)^{-2}. \quad \dots\dots(125)$$

The change which will result behind surface $(i+1)$ must now be determined. From equation (114) we have

$$\frac{\partial\sigma'_{i+1}}{\partial\sigma_{i+1}} = n_{i+1}, \quad \dots\dots(126)$$

whence

$$\frac{\partial\sigma'_{i+1}}{\partial\sigma'_i} = n_{i+1} \frac{\partial\sigma_{i+1}}{\partial\sigma'_i},$$

and extending this expression surface by surface we obtain

$$\frac{\partial\sigma'_k}{\partial\sigma'_i} = \frac{\partial\sigma_{i+1}}{\partial\sigma'_i} n_{i+1} \frac{\partial\sigma_{i+2}}{\partial\sigma'_{i+1}} n_{i+2} \dots\dots \frac{\partial\sigma_k}{\partial\sigma'_{k-1}} n_k, \quad \dots\dots(127)$$

and the relation between these transfer coefficients at successive surfaces is, therefore,

$$\frac{\partial\sigma'_k}{\partial\sigma'_i} = \frac{\partial\sigma_{i+1}}{\partial\sigma'_i} n_{i+1} \cdot \frac{\partial\sigma'_k}{\partial\sigma'_{i+1}}. \quad \dots\dots(128)$$

We now have the complete framework for the computation of the transfer coefficients. Equation (125) is used first to calculate $\partial\sigma_{i+1}/\partial\sigma'_i$ for each surface except the last. Then at the last surface $\partial\sigma'_k/\partial\sigma'_k = 1$, and equation (128) is used to calculate $\partial\sigma'_k/\partial\sigma'_{k-1}$, and the computation may be continued surface by surface through the system. Finally these basic transfer coefficients are used in conjunction with the single surface coefficients, giving expressions of the form

$$\frac{\partial\sigma'_k}{\partial c_i} = \frac{\partial\sigma'_k}{\partial\sigma'_i} \frac{\partial\sigma'_i}{\partial c_i},$$

and returning to the quantity s' instead of its reciprocal, we have

$$\frac{\partial s'_k}{\partial c_i} = -s'_k \frac{\partial\sigma'_k}{\partial\sigma'_i} \frac{\partial\sigma'_i}{\partial c_i}, \quad \dots\dots(129)$$

with corresponding expressions for $\partial s'_k/\partial d_i$ and $\partial s'_k/\partial n_i$.

§ 5. A NUMERICAL EXAMPLE

In the course of correcting a certain photographic objective having ten surfaces, a change of curvature of 0.001 mm.⁻¹ was made at the seventh surface, the radius of curvature of that surface changing from 1091 mm. to 521.78 mm. In a pencil of approximately 17° obliquity the following results were obtained.

$$\frac{\partial s_k'}{\partial c_7} = -2112.5, \quad \frac{\partial t_k'}{\partial c_7} = -2416.0,$$

and hence

$$ds_k' = -2.113, \quad dt_k' = -2.416,$$

giving

$$\text{new } s_k' = 47.273, \quad \text{new } t_k' = 47.661.$$

Using the alternative relations (108) to (110) there was obtained

$$dt_k' = -2.335, \\ \text{new } t_k' = 47.742.$$

A full trace of the altered system gave the final positions of the two foci as

$$s_k' = 47.338, \quad t_k' = 47.801,$$

which represents an accuracy of about 3%.

A TRANSFER METHOD FOR DERIVING THE EFFECT ON THE IMAGE FORMED BY AN OPTICAL SYSTEM FROM RAY CHANGES PRODUCED AT A GIVEN SURFACE

By A. L. M'AULAY,
University of Tasmania

MS. received 25 January 1944

ABSTRACT. A transfer method is described depending on the invariance of a function of the position change of successive foci of a narrow pencil as it passes through an optical system.

The method described is approximate, as the invariant is applied to wide pencils, but by reason of its speed it is a useful tool in design.

§ 1 INTRODUCTION

IN a paper by M'Aulay and Cruickshank (1945) a method is described by which the effect on the final image formed by an optical system may be computed when alterations are made at any surface of the system. The method naturally breaks into two parts. In the first part the changes which rays suffer as they pass through the surface under consideration are evaluated, and in the second part these changes are transferred to the final image formed by the system.

The present paper describes another method of transferring the changes produced at any surface to the final image. It is based on a theorem of a form similar to Lagrange's, but is applicable to pencils of any obliquity.* If the theorem is applied to narrow pencils it is exact for differential changes, but in the use here described it is applied to wide pencils and is approximate. It has been used principally as a primary computing method where speed is more important than exactness, and it has been found of considerable practical value. It is believed that its potentialities are greater than the particular development discussed.

§2 AN INVARIANT OF THE LAGRANGE FORM FOR OBLIQUE PENCILS

The method deals essentially with pencils, not with rays. It follows the change in position of the successive foci of a pencil which suffers a small change in position due to an alteration made in the constants of one surface of the system through which it is passing. Figure 1 shows a narrow pencil ABC focusing

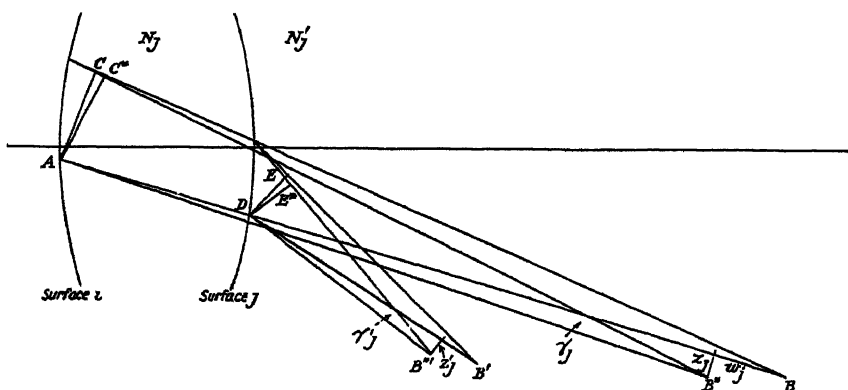


Figure 1.

at B after passing through the i th surface of the system. An alteration made to the i th surface causes the pencil to focus instead at B.* The co-ordinates of the change are w , and z , measured along and at right angles to the original axis of the pencil. They are given the positive sign when they have the direction shown in the figure, as this relates them conveniently to the usual sign of U , the angle under an incident ray. If γ_j is the angle between the extreme rays of the pencil and N_j is the refractive index of the medium between the i th and the j th surfaces, it may be shown that

- (i) $N_j \gamma z$ remains unchanged as the pencil is followed through successive surfaces in its original and changed form, and
- (ii) $N_j \gamma^2 w$ also remains unchanged.

The second relation holds, however, only if the change is such that a spherical wave remains spherical after the change. It is a useful relation for some purposes, but will not be further considered here.

* This theorem, given in §2, was discovered by the author. The referee has pointed out that it was known independently (see, e.g., v. Rohr, *Formation of Images in Optical Instruments*, 1920, p. 174), and no proof is therefore given here.

It is important to note that the change of a ray is not being followed. The zs are co-ordinates of the changes in the successive focal points of a pencil, but the actual rays which have been followed in the above argument change from surface to surface. Unless this point is kept clearly in mind it is easy to make serious mistakes when the principle is used as a computing tool

Alternative statement of the principle. Figure 2 shows a pencil with its axis

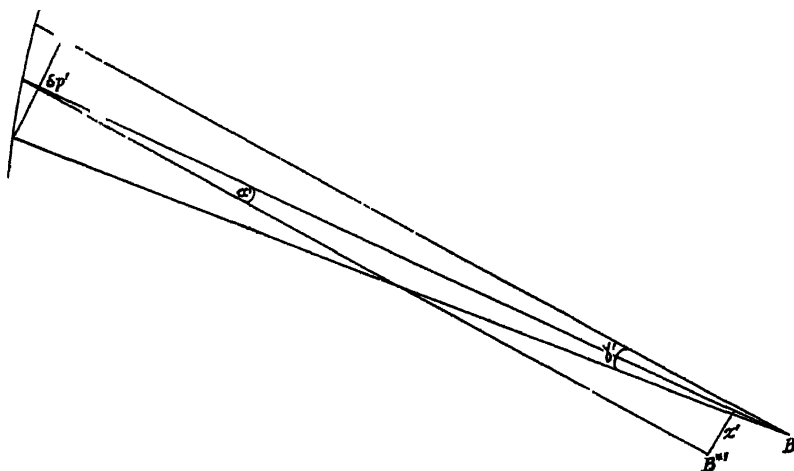


Figure 2

displaced a small distance z' . It is evident that if α' is the angle through which the axis is displaced,

$$N'\gamma'z' : N'\frac{\delta p'}{S'}z' = N'\alpha'\delta p',$$

where S' is the length of the axis of the pencil from surface to focus, $\delta p'$ is the width of the pencil at the surface and α' the angle turned through by the axis of the pencil due to the alteration. This form is often best suited for obtaining the value of the invariant at an altered surface.

§3. APPLICATION OF THE PRINCIPLE TO THE TRANSFER OF DIFFERENTIAL CHANGES PRODUCED AT THE i TH SURFACE OF A SYSTEM

In the paper by M'Aulay and Cruickshank quoted above, a change in the aberrations of an optical system due to an alteration of the i th surface was discussed. The system is primarily analysed by considering six rays, three axial (marginal, zonal and paraxial) and three oblique (a , pr and b). The same rays are considered in the method to be described, but they are treated as the axes of six pencils. It is shown in the paper quoted that the change in angle under a ray refracted at a spherical surface can be given as a function of the curvature or refractive-index change. In what follows, it will be assumed that such a surface alteration has produced changes dU' in the angles under the refracted rays, and that these have been computed.

The transfer theorem is used to convert these dU to displacements z' perpendicular to the directions of the changed rays in the neighbourhood of the final image plane. It is evident that z' is very nearly the co-ordinate of the displacement of the focus of the narrow pencil which in the notation of the last section would have been written z_k' , where k is the final surface.

Approximate evaluation of z' . The theorem is used in the form

$N'\alpha'\delta p'$ is invariant through the system.

This invariant is found for the i th surface and used to obtain z' after refraction at the last surface by the relation

$$N_i'\alpha_i'\delta p_i' = N_k'\alpha_k'\delta p_k' = N_k'\frac{z'}{S_k'}\delta p_k',$$

or

$$z' = N_i'\alpha_i'\frac{\delta p_i'}{\delta p_k'}S_k',$$

where the refractive index after the last surface considered is unity.

N_i' is available from the original data, α' is the change dU_i' which has been supposed computed, and S_k' , the distance measured along the ray from the last surface to the final image, can easily be obtained from the trace. It remains to evaluate $\delta p_i'/\delta p_k'$.

It is here that the most serious approximation is made. The object is to find sufficiently good values of $\delta p'$ from the basic six-ray trace without much extra

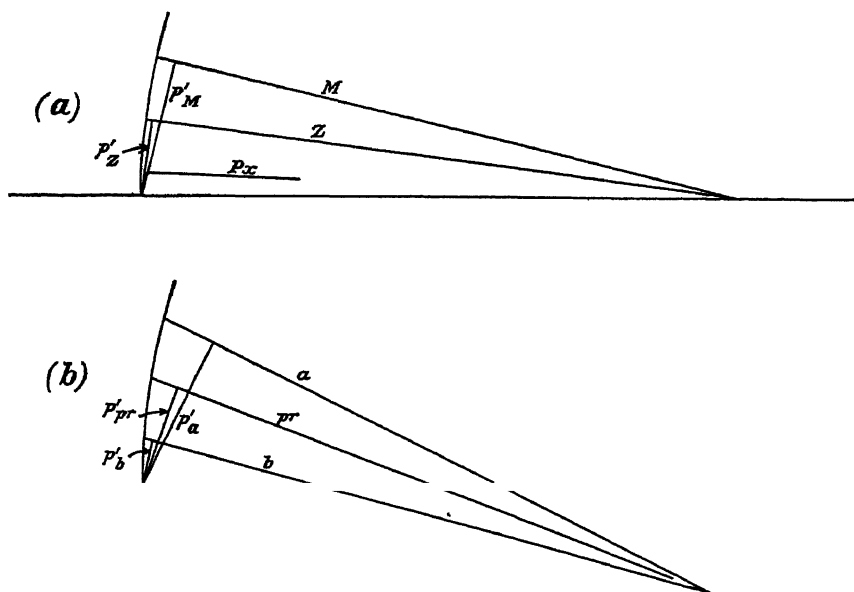


Figure 3

computation. In the tracing method used, the value of $L'\sin U' = p'$ is, incidentally, found for each ray. This is the perpendicular distance from the pole of a surface to the ray under consideration.

Figure 3 (a) shows marginal, zonal and paraxial rays, figure 3 (b), a , pr and b rays and the perpendiculars dropped from the pole on them. The following

values are taken for $(\delta p'/\delta p'_k)$:

$$\begin{aligned}\delta p'_x/\delta p'_{xk} &= p'_x/p'_{xk}, \\ \delta p'_z/\delta p'_{zk} &= p'_M/p'_{Mk}, \\ \delta p'_M/\delta p'_{Mk} &= 2(p'_M - p'_Z)/(p'_M - p'_Z)_k - p'_Z/p'_{Zk}, \\ \delta p'_{pr}/\delta p'_{prk} &= (p'_a - p'_b)/(p'_a - p'_b)_k, \\ \delta p'_a/\delta p'_{ak} &= (p'_a - p'_{pr})/(p'_a - p'_{pr})_k \\ &\quad + \frac{1}{2}[(p'_a - p'_{pr})/(p'_a - p'_{pr})_k - (p'_{pr} - p'_b)/(p'_{pr} - p'_b)_k], \\ \delta p'_b/\delta p'_{bk} &= (p'_{pr} - p'_b)/(p'_{pr} - p'_b)_k \\ &\quad - \frac{1}{2}[(p'_a - p'_{pr})/(p'_a - p'_{pr})_k - (p'_{pr} - p'_b)/(p'_{pr} - p'_b)_k].\end{aligned}$$

It is evident that other extrapolations for the marginal, a and b rays might be used. The rather crude forms given are easy to compute and reasonably good.

§ 4 THE ABERRATION CHANGES FROM THE z

Spherical aberration and curvature. The aberration changes can be stated simply in terms of z' . The changes in spherical aberration and tangential curvature of field will be written down without further discussion as their derivation follows the same lines as that given by Cruickshank, Part I (1945).

For spherical aberration

$$dL'_M = dl' - dL'_M = z'_x/u' - z'_M \operatorname{cosec} U'_M,$$

with an exactly similar expression for zonal spherical aberration.

For curvature

$$dX'_T = (z'_a \cos U'_b - z'_b \cos U'_a) \operatorname{cosec} (U'_a - U'_b) - z'_x \operatorname{cosec} U'_x.$$

The notation is that of Conrady's *Applied Optics and Optical Design* or the papers quoted above. A change of sign, however, has been introduced for X'_T , and the signs for coma and distortion will also be changed. It has been found convenient in this work to adopt a positive sign for an aberration that arises when parallel pencils of small inclination to the principal axis pass from air through a convex glass surface.

Coma and distortion. The co-ordinate changes perpendicular to the principal rays can be used direct to estimate these aberrations. The change in coma is sufficiently well given by the change in distance between the principal ray and the median of the a and b rays. It is given by

$$d(\text{coma}') = z'_{pr} - \frac{1}{2}(z'_a - z'_b) \sec \frac{1}{2}(U'_a - U'_b).$$

The treatment of change of distortion, involving as it does an imaginary ray not included in the trace, requires a little more discussion. An ideal ray is considered which is identical with the principal ray till it strikes the surface at which an alteration takes place. Subsequently this ray is treated as though it were refracted according to the paraxial form of the computing equations. The change in angle under the ideal ray after refraction at the altered surface therefore differs for that obtained for the principal ray. This change of angle is computed and is transferred to a z' change in the neighbourhood of the final image, which is written z'_{prx} and is given by

$$z'_{prx} = N' \alpha'_{prx} (p'_x/p'_{xk}) l'_k.$$

The change in distortion is now given in the same manner as the change in coma by

$$d(\text{dist}') = z'_{prx} - z'_{pr}$$

REFERENCES

- M'ALLAN, A. L. and CRICKSHANK, F. D., 1945 *Proc. Phys. Soc.* **57**, 302
 CRICKSHANK, F. D., 1945. *Proc. Phys. Soc.* **57**, 350

THE RELATION BETWEEN THE BRIGHTNESS AND TEMPERATURE OF A TOTAL RADIATOR

By E. F. CALDIN,
 Queen's College, Oxford

MS received 31 January 1945, in revised form 14 June 1945

ABSTRACT Values are calculated for the brightness of a total radiator at temperatures between 2000° K and 3120° K., using the standard C.I.E. relative luminosity curve and the Planck radiation law. A simple logarithmic equation is found to express the relation between brightness and temperature to within $\pm 1\%$ over this temperature range. The ratio of the new international candle to the lightwatt is calculated from the results.

§1 INTRODUCTION

IT is sometimes desirable to use an expression for the brightness of a total radiator as a function of temperature. From the literature it appears that nothing has been published on the subject since the adoption of the C.I.E. relative luminosity curve in 1924. Hyde, Forsythe and Cady (1919) published some values of the brightness for the range 1700° K. to 2650° K.; they used their own values for the relative luminosity function, however, and these differ from the standard values later adopted. They expressed the results in terms of a rather complex four-term equation. Nernst (1906) expressed the relation experimentally found between the brightness (B) and temperature (T) of a total radiator for the range 1460° K. to 2280° K. by the following equation:—

$$\log B = -A/T + C,$$

but the relative luminosity curve for the observer concerned is not recorded.

Values for the brightness of a total radiator have been calculated from the C.I.E. relative luminosity data and the Planck equation for fifteen temperatures ranging from 2000° K. to 3120° K. The following approximate formula, which has the same form as Nernst's relation, has been obtained for the brightness of a total radiator over this temperature-range:—

$$\log_{10} B = 4.275 - 10860/T, \quad \dots\dots(1)$$

where B is the brightness of the total radiator (lightwatts cm.⁻²) and T is its temperature (°K.).

§ 2 CALCULATION OF BRIGHTNESS

The brightness of a total radiator at a given temperature is calculated from the equation

$$B = \frac{1}{\pi} \int_0^{\infty} J_{\lambda} V_{\lambda} d\lambda, \quad \dots\dots (2)$$

where B is the brightness (lightwatts cm^{-2}). $J_{\lambda} d\lambda$ is the total energy flux per unit area from a total radiator (watts cm^{-2}) between the wave-lengths $\lambda \pm \frac{d\lambda}{2}$ (microns), and V_{λ} is the value of the relative luminosity factor at the wave-length λ , as given by the 1924 C.I.E. standard luminosity curve (Judd, 1931). The factor $\frac{1}{\pi}$ appears because the brightness of an element of surface, i.e. the ratio of intensity to projected area in a given direction, bears the ratio $\frac{1}{\pi}$ to the total luminous flux per unit area from the same element of surface (assuming that the surface has the properties of a perfect diffuser). In the present case the expression $\int_0^{\infty} J_{\lambda} V_{\lambda} d\lambda$ represents the total luminous flux per unit area of surface, since it is derived from $J_{\lambda} d\lambda$, which represents the total energy flux per unit area for a small wave-length range. Hence the brightness is given by $\frac{1}{\pi} \int_0^{\infty} J_{\lambda} V_{\lambda} d\lambda$.

The values of $J_{\lambda} d\lambda$ are derived from Planck's formula,

$$J_{\lambda} d\lambda = \frac{C_1 \lambda^{-5} d\lambda}{e^{C_2/\lambda T} - 1}, \quad \dots\dots (3)$$

where C_1 and C_2 are constants and $J_{\lambda} d\lambda$, λ , and T have been previously defined

Convenient tables for deducing values of B from equation (2) are those of Skogland (1929). These tables employ the values for C_1 and C_2 given in the International Critical Tables, namely $C_1 = 3.703 \times 10^{-5} \text{ erg sec}^{-1} \text{ cm}^{-2}$, $C_2 = 14330 \text{ micron degrees}$. The range of temperature considered is 2000° K . to 3120° K .; this is covered in steps of 20° in table 1 and 80° in table 2. Table 1 gives at each temperature, besides values of $J_{\lambda}/J_{0.59}$ for a series of wave-lengths, values of $J_{0.59}/C_1$ to 5 significant figures (units $\cdot \text{microns}^{-5}$). From these latter values, taking $C_1 = 3.703 \times 10^{-5} \text{ erg sec}^{-1} \text{ cm}^{-2}$ and $V_{0.59} = 0.7570$ (Judd, 1931), we calculate the value of $J_{0.59} V_{0.59}$ at a series of temperatures, at intervals of 80° , from 2000° K to 3120° K . Skogland's table 2 gives at each of these temperatures, for a series of wave-lengths from 0.40μ to 0.76μ in steps of 0.01μ , values for $J_{\lambda} V_{\lambda}/(JV)_{\text{max}}$, where $(JV)_{\text{max}}$ is the maximum value of $J_{\lambda} V_{\lambda}$ at the given temperature. Reading off the values for $J_{0.59} V_{0.59}/(JV)_{\text{max}}$, and substituting the value already obtained for $J_{0.59} V_{0.59}$, we find $(JV)_{\text{max}}$ for the given temperature. Table 2 gives also the sum of the values of $J_{\lambda} V_{\lambda}/(JV)_{\text{max}}$, at each temperature, i.e.,

$$\sum_{\lambda=0.40 \mu}^{\lambda=0.76 \mu} J_{\lambda} V_{\lambda}/(JV)_{\text{max}},$$

the summation referring to steps of 0.01μ . The values at the wave-length limits

are not more than about 0.03% of the maximum values, and the sum does not differ by more than 1 in 10^4 from

$$\sum_{\lambda=0}^{\lambda=\infty} J_{\lambda} V_{\lambda} / (JV)_{\max}.$$

Multiplying the sum by 0.01, we obtain, therefore, a value (in microns) of

$$\int_0^{\infty} [J_{\lambda} V_{\lambda} / (JV)_{\max}] d\lambda, \quad \text{i.e.} \quad \frac{1}{(JV)_{\max}} \int_0^{\infty} J_{\lambda} V_{\lambda} d\lambda,$$

which is accurate at least within 1 in 10^4 . Multiplying this by the value of $(JV)_{\max}$ already found, we obtain the value of $\int_0^{\infty} J_{\lambda} V_{\lambda} d\lambda$; and hence, dividing by π , the brightness of a total radiator, at the given temperature

§3 RESULTS AND DISCUSSION

In the table below are given values for the brightness B , in lightwatts per sq. cm., and for $\log_{10} B$, at a series of temperatures from 2000°K. to 3120°K.

The value of $\log_{10} B$ is found to vary nearly linearly with $1/T$, and the following formula has been tested :—

$$\log_{10} B = 4.2750 - 10860 \, 0/T. \quad \dots\dots(1)$$

Values derived from this formula are given in the table and compared with the values directly calculated.

Temperature ($^{\circ}\text{K.}$)	Brightness B (lightwatts cm^{-2})	$\log_{10} B$ (a)	$\log_{10} B$ calculated from eqn. (1) (b)	Difference (a)–(b)
2000	0.0706	2.8490	2.8450	+0.0040
2080	0.1137	2.0556	2.0538	+0.0018
2160	0.1767	2.2473	2.2472	+0.0001
2240	0.2665	2.4257	2.4268	–0.0011
2320	0.3910	2.5922	2.5940	–0.0018
2400	0.5595	2.7478	2.7500	–0.0022
2480	0.7827	2.8936	2.8959	–0.0023
2560	1.0730	0.0306	0.0328	–0.0022
2640	1.443	0.1594	0.1613	–0.0019
2720	1.910	0.2809	0.2823	–0.0014
2800	2.486	0.3956	0.3964	–0.0008
2880	3.192	0.5040	0.5042	–0.0002
2960	4.040	0.6068	0.6061	+0.0007
3040	5.060	0.7042	0.7026	+0.0016
3120	6.262	0.7967	0.7942	+0.0025

It is evident that the values of B are given within $\pm 1\%$ (an accuracy sufficient for many purposes) by the empirical equation (1).

The effect on the formula (1) of using other values for C_1 and C_2 may be noted. A change in C_1 produces a proportionate change in all the values of $J_{\lambda} d\lambda$, and hence in $\int_0^{\infty} J_{\lambda} V_{\lambda} d\lambda$, and so can be met by changing the constant 4.275 to

$[4.275 + \log_{10}(C_1/3.703)]$, where C_1 is expressed in $\text{erg sec.}^{-1} \text{cm.}^2 \times 10^{-5}$. A change in C_2 affects the temperature-dependent term; since $\exp(C_2/\lambda T) \gg 1$ for the ranges of temperature and wave-length concerned, it is only necessary to change the constant 10860 to $(10860 \times C_2/14330)$, where C_2 is expressed in micron degrees. Thus if we adopt the recently suggested values of Birge (1941), namely $C_1 = 3.7430 \times 10^{-5} \text{ erg sec.}^{-1} \text{cm.}^2$ and $C_2 = 14384.8 \text{ micron degrees}$, the relation becomes

$$\log_{10} B = 4.279 - 10902/T.$$

§ 4. PHOTOMETRIC UNITS. THE RATIO OF THE NEW INTERNATIONAL CANDLE TO THE LIGHTWATT

It is possible to convert the above values of brightness directly from lightwatts per sq. cm. to "new international candles" per sq. cm., and to calculate the ratio of the new candle to the lightwatt, because the new candle is defined in terms of the intensity of a total radiator at a fixed temperature. This is of interest because the new candle is likely to come into force in the near future as the international unit of luminous intensity. By definition, the luminous intensity of a total radiator at the temperature of solidification of platinum is 60 new candles (International Lighting Vocabulary, C.I.E., 1938). The brightness in lightwatts per cm^2 of a total radiator at this temperature (taken here as 2045.9°K . (Wensel, 1937) is found, by interpolation from the directly calculated figures in the table above, to be 0.0932. The ratio of the new candle to the lightwatt is thus $60/0.0932 = 643$ for the values of C_1 and C_2 used in computing the figures in the table; and to convert the values given above for the brightness of a total radiator to new candles per cm^2 it is only necessary to multiply each value by 643, or to add 2.808 to its logarithm.

ACKNOWLEDGMENT

The author's thanks are due to Mr. J. S. Preston, of the National Physical Laboratory, for his advice and interest in this work.

Note added in proof. Mention should be made of the work of Ives (1926), who calculated values of the brightness of a total radiator at temperatures from 1200°K . to 10000°K . In these calculations, an approximate formula was used to express the C.I.E. relative-luminosity data, and errors of the order of 1% are likely. The author was unaware of this work when the above paper was written. He has also been informed of a paper by W. Geiss (*Licht*, 12, 33 (1943)), but has not been able to see a copy.

REFERENCES

- BIRGE, R. T., 1941 *Rev. Mod. Phys.* 13, 233.
 HYDE, E. P., FORSYTHE, W. E. and CADY, F. E., 1919. *Phys. Rev.* 13, 45.
 IVES, H. E., 1926 *J. Opt. Soc. Amer.* 12, 75.
 JUDD, D. B., 1931. *Bur. Stand. J. Res., Wash.*, 6, 465.
 NERNST, W., 1906. *Phys. Z.* 7, 380.
 SKOGLAND, J. F., 1929. *Misc. Pub. Bur. Stand., Wash.*, no. 86.
 WENSEL, H. T., 1937. *J. Res. Bur. Stand.* 6, 1119.

REVIEWS OF BOOKS

Waveform Analysis, by R. G. MANLEY. Pp. vii+275 and 3 plates. (London: Chapman and Hall, Ltd., 1945.) 21s.

This book is not a treatise on harmonic analysis in the conventional sense but, as the author correctly emphasizes in the sub-title and preface, a practical guide to the rapid interpretation of wave records. Its primary aim is to explain, against a background of more formal analysis, methods, involving merely inspection and simple measurement, for estimating the frequency, amplitude and phase of the principal harmonic components of a large class of complex wave-forms which occur especially in vibration engineering. There are ten chapters, entitled: I, *Sine waves in combination*; II, *General properties of harmonic series*; III, *Basic analysis of recorded waveforms*; IV, *The envelope method*; V, *Method of superposition*; VI, *Fourier series mathematical analysis*; VII, *Numerical methods*; VIII, *Mechanical and other aids to analysis*; IX, *Practical requirements for waveforms*; X, *Lissajous figures*.

Chapter I illustrates graphically and trigonometrically the synthesis of wave patterns (especially beats) from two or more sine components, and, together with Chapter II, provides that foundation of experience of the anatomy of wave-forms which is necessary for a ready diagnosis of harmonic content.

Chapter III seems a little dilute, and might be made more concise with advantage, but it serves to pave the way to the actual measuring of records and hence to Chapter IV, which forms the pivot of the book. This chapter, essentially diagnostic in character, shows how the construction of "envelopes", to touch the crests and troughs (or, more generally, to pass through any repeated characteristic kinks) of the undulations, makes it possible to identify the harmonics and to determine their characteristics. A lucid explanation is given of the surprising amount of information which may be extracted from a wave-form by an experienced interpreter without recourse to elaborate analysis. The method appears to have especial application to the routine examination of vibrograph, strain-gauge and similar records which are nowadays obtained in certain industries in numbers too great to permit of detailed analysis. A systematic procedure for applying the envelope method is developed in a number of tables somewhat reminiscent of chemical analysis.

For certain frequency ratios amongst the harmonics (such as 3 2 1), the envelope method fails. Chapter V shows how it is possible to meet these cases by dividing the complete cycle of the given wave into a number of equal parts which, by suitable summation and differencing, permit partial, and in favourable circumstances complete, separation of the harmonics.

It may be remarked that the envelope technique and similar simple methods, while of great service in resolving a wave composed of relatively few harmonics of comparable amplitude, do not promise so much help to the physicist, meteorologist and economist, who will still have to resort to Fourier, periodogram and correlogram analysis when confronted, as they frequently are, by waves having a long retinue of progressively diminishing harmonics.

The theoretical Chapter VI treats the formal representation of functions (including the usual examples of square and saw-tooth profiles) by Fourier series, and gives an adequate sketch of the mathematical background. While very properly leaving full rigour to the mathematicians, it does somewhat more than merely give plausible derivations of formulae. One cannot help regretting, however, that the nature of the book scarcely permits presentation of the connection between the author's "periodic existence functions" and their analogues, the Cauchy discontinuous functions. Gibbs' phenomenon at discontinuities is treated briefly, and Fourier integrals are mentioned but not applied.

Chapter VII gives a clear account of the numerical Fourier analysis of waves specified by equidistant ordinates, including computation schedules for schemes using 24 and 48 ordinates.

Chapter VIII describes in detail a versatile drawing-table type of mechanical harmonic analyser, and briefly mentions alternative devices, including electrical filter circuits. No attempt is made to survey the very rich field of mechanical and electrical aids which now

General comments on the practical problems of recording and allowing for response distortion are given in Chapter IX, and the book closes with a chapter on the characteristics of two vibrations compared by means of Lissajous figures.

There are several appendices dealing with mathematical points, a list of elementary trigonometrical formulae, a useful table of sines and cosines adapted to facilitate harmonic analysis, a glossary of terms, a bibliography and an index.

A minor criticism of the presentation is that elementary details of mathematical manipulation are occasionally given more fully than is necessary if the reader is sufficiently mature to appreciate the chapter on Fourier analysis, but this may encourage further reading, and it would therefore be beneficial if the bibliography were somewhat enlarged. Viewed as a whole, the book is well written and competently composed, and should be of considerable value to those concerned with vibration engineering. There are few misprints, of which the most obtrusive is the insertion of the terminal "s" of Lissajous in L'Hopital. The printing and binding are quite good.

M. S. JONES.

Introduction à l'étude des champs physiques, by J. FRANIER. Pp. iv + 251. (Paris: Dunod, 1941.)

This book is not an introduction to *field theories* as now understood, but a general treatment of steady-state boundary problems where the differential equation is that of Laplace or of Poisson, that is of problems in electrostatics, magnetism, heat conduction, capillarity or hydrodynamics. The treatment is relatively elementary, and is confined in the main to problems in two dimensions, or those with spherical symmetry, so that there is no call to introduce Legendre functions, spherical harmonics or Bessel functions. The French nation has given the world some of its greatest geometers, and the tradition of geometrical reasoning has always been strong there, so that it is not surprising to note that, in this book, transformations from one co-ordinate system to another are seen from a rather more directly geometric aspect than they usually are in books written by Englishmen, who tend to see them as algebraic transformations.

After a preliminary chapter in which conjugate complex functions and systems of orthogonal curves are considered, the book is divided into chapters, each dealing with some phenomenon in which Laplacian fields present themselves, beginning with electrostatics. By taking the potential due to a point charge, and that due to an infinite straight line (a point charge in two dimensions), other cases are built up, the most elegant being that of a circular cylinder surrounding two line charges of opposite sign, a case which is found when considering the case of four line charges, by noting that there is one equipotential which is circular. A few cases are solved by means of conjugate functions, but these are generally introduced experimentally "to find out what cases they will represent", rather than deliberately to solve particular problems posed beforehand. Even after the theorem of Schwartz on the behaviour of functions at branch points has been introduced, little use is made of it, though in fact it enables us to select with certainty the transformation appropriate to any plane problem with rectilinear boundaries. The chapter on magnetism finds its interest of course in doublets and other multiple charges, and does not carry the matter very far. The field due to a sphere in a field which would otherwise be uniform is worked out, but the case, so dear to most text-book writers, of an ellipsoid in a uniform field, is omitted. The misprints in the middle of page 121, where R^3 is omitted from the denominator of two successive equations, might mislead beginners. In the section on hydrodynamics we have such problems as the flow past a cylinder, and attention is then turned to cases where there is a free stream-line, as when a jet issues through a slit. No use is made here of the method devised by Kirchhoff of introducing a subsidiary plane in which the co-ordinates are the absolute magnitude and the direction of the velocity, but a great deal is extracted from the simple transformation, attributed to Joukowski, $x = \frac{1}{2}(r + r^{-1}) \cos \theta$ and $y = \frac{1}{2}(r - r^{-1}) \sin \theta$, which analysts would think of as $w = (z + 1/z)$. This, with its inverse transformation, and the addition of different fields

of flow, yields an astonishing amount of information about flow round obstacles of practical interest, such as aerofoils

The last quarter of the book is of a rather different nature, and considers the use of experimental methods for solving complicated problems—measurements of electrical resistance of diagrams cut out of metal foil, or tracing potential lines by means of probe measurements in an electrolyte, or plotting out a field by means of iron filings above a magnet, for example. The author has clearly had a very considerable experience with these methods, and is able to give advice on experimental arrangements which must be of immense value to any who propose to use the methods. Among them he includes the graphical method in which a network of orthogonal curves is drawn freehand, so spaced as to divide the field approximately into squares.

There is no doubt that the book will be of value to many, in showing the essential unity of the diverse branches of physics with which it deals, in enabling them to devise methods for attacking problems with which they are confronted, and as a book of reference, giving the solutions of many problems. When we consider that it was published in France in 1941, we can only imagine the difficulties which must have confronted author and publisher, and must admire them for producing the book at all, that they produced such a good one must almost be a matter of wonder.

J. H. A.

The Quantum Theory of Radiation, by W. HEITLER. Second Edition. Pp. viii + 272. (London and Oxford: Sir Humphrey Milford, 1944.) 20s. net.

Ten years have passed since the first edition of this book appeared, and it has clearly established itself in the interval as an authoritative treatment of its extremely difficult subject. Although progress, particularly on the experimental side, has been steady, there has not yet been any spectacular advance which would call for a revision and re-writing of the theory, and in fact the present volume does not differ essentially from the first edition, save for interpolations here and there and an addition to the appendix.

One day, of course, a complete change of approach will be found, which will at one stroke give us a quantum theory completely in harmony with the postulates of relativity, and which will avoid the difficulties we find at present in connexion with the infinite self-energy of the point electron. Till this new formulation is made, it is difficult to imagine any more useful, or more instructive, text-book than this, though we may at times wish that someone would write a simple text to serve as an introduction to the subject.

J. H. A.

Tables of Elementary Functions, by F. EMDE. Pp. xii + 181. (Leipzig: Teubner, 1940; photographic reprint published under licence by J. W. Edwards, Ann Arbor, Mich., U.S.A., 1945.) \$3.20.

Sechsstellige trigonometrische Tafeln, by H. BRANDENBURG. Pp. xxiv + 304. (Leipzig: Lorentz, 1932; photographic reprint published under licence by J. W. Edwards, Ann Arbor, Mich., U.S.A., 1945.) \$5.00.

It is regretted that, in the reviews of these two books in the July issue of the *Proceedings* (this volume, pp. 368–370), no mention was made of the fact that the sole agent for these tables in Great Britain is *The Scientific Computing Service, Ltd.*, of 23 Bedford Square, London W.C. 1.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 57, PART 6

November 1945

No. 324

THE ELECTROPHYSIOLOGICAL ANALYSIS OF THE FUNDAMENTAL PROBLEM OF COLOUR RECEPTION

BY RAGNAR GRANIT,

Director of the Nobel Institute for Neurophysiology,
The Royal Caroline Institute, Stockholm

Fourteenth Thomas Young Oration, delivered 29 June 1945

§ 1. INTRODUCTION

IN a memorable lecture at the Royal Society in 1801 Thomas Young expressed his well-known views on the mechanism of colour reception. This was done almost in passing, and accompanied by the comment that he regarded them as something of a corollary to Newton's great discovery. It was impossible, he said, to conceive of the many spectral colours as being represented by an equally great number of optic nerve fibres of different type. For this reason he suggested that the number is limited. He fixed the number of fibre types at three, but this limitation is less essential than the three main generalizations implied in his thesis: (i) that the analysing mechanism is placed in the peripheral visual apparatus, (ii) that the number of colour-sensitive elements is relatively limited, (iii) that these elements represent widely different regions of the visible spectrum. Certain fundamental aspects of the problems contained in these assumptions can now be investigated with the aid of very direct methods. It has become possible to record the impulses in single fibres of the optic nerve of animals.

It is well known to most physicists that Professor Adrian in this country first showed that all sense organs respond to stimulation by discharging trains of impulses of constant size through their nerve fibres. The electrical nature of the nerve impulse as a conducted wave of negativity has been known since the work of Du Bois Raymond in 1849, but by introducing electrical amplification and by cutting down the nerves from certain receptors until only single fibres were left, Adrian succeeded in demonstrating that the response to variations in stimulus intensity is a variation in frequency of an impulse of constant size in the individual fibres (Adrian, 1932). This result is the basis of modern physiological analysis of all sensory mechanisms. The retina was also included in Adrian's pioneer programme of research (1927), and several important results were obtained. But at that time, in the late twenties, the new technique opened up so many prospects to physiology that the seemingly hopeless and formidable problem of isolating single fibres among the several hundred thousands of the optic nerve was left unsolved.

The American physiologist Hartline (1938) solved this technical problem for the first time by a very painstaking method of microdissection. He used the excised eye of the frog, removed lens and cornea and lifted up fibres from the retina on to his electrode in the region where they converge to form the blind spot. This technique is extremely laborious, and I remember Dr. Hartline telling me once that every eighth experiment succeeded, provided, of course, that the experimenter had acquired sufficient skill to succeed at all. (For a review of Hartline's work, see *J. opt. Soc. Amer.* 1940.)

§ 2. MEASUREMENTS ON ISOLATED FIBRES

Our own microelectrode technique is a great deal simpler. For work on colour reception it is essential that the technique for isolation of fibres should not be too difficult. It must be applicable not only to excised eyes but also to mammalian eyes in the living narcotized or decerebrated animal. The microelectrode is a thin platinum wire, isolated with glass down to the tip. It is applied to the inside of an eye from which cornea and lens have been removed. The microelectrode touches the fibres running along the retina to the blind spot. A schematic picture is shown in figure 1, referring to the rat's retina. This picture

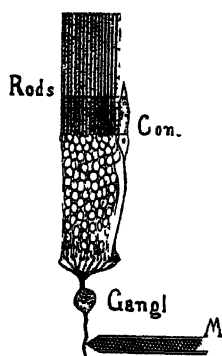


Figure 1. Diagram illustrating microelectrode on optic nerve fibre of rat. Several receptors and bipolars converge to form the unit recorded from. (Granit, *Acta Physiol Scand* 2, 1941 a.)

also illustrates the fact that several receptors converge towards the single fibre isolated. There are few cones in the rat's retina. One cone and a much greater number of rods make up the convergence unit assumed to be isolated by the microelectrode. Such isolated fibres have very different properties. The records in figure 2 illustrate the simplest type: the isolated unit responds to an increase in intensity from the threshold upwards, with an increasing number of impulses at an increasing rate of discharge. These particular records are from the eye of a guinea pig in which such units form the great majority; it was maximally sensitive to green light around $530\text{ m}\mu$, as shown by figure 3. The ordinate of this curve is the inverse value of the relative quantum intensity necessary to elicit one impulse, as in the uppermost record of figure 2. The animal was light-adapted.

This experiment illustrates the main principles of the microelectrode experiment on colour reception. It is a technique for measuring the absolute threshold

to spectral light. In actual practice such experiments are not carried out photographically, but the impulses, or, as we say, the *spikes*, are amplified and led to a loud-speaker. The spectral energy is increased by means of a calibrated wedge until a sharp report is heard in the loud-speaker. Single fibres are not always

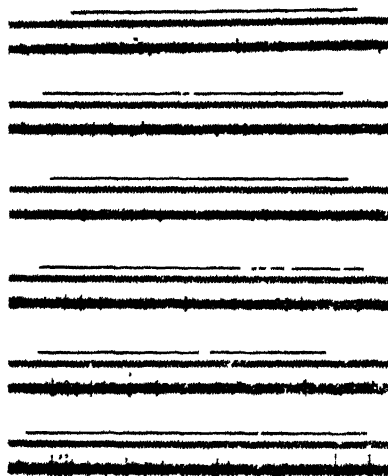


Figure 2 Guinea pig Isolated unit with maximal sensitivity in $530\text{ m}\mu$ responding to different intensities, increasing from record to record downwards. Uppermost line in each record = light signal. Time in $1/50$ sec. between this and spike records. Uppermost record at threshold strength, lowermost record at 10.7 times threshold energy (Granit, *Acta Physiol Scand.* 3, 1942.)

obtained, and quite often one has to be content with a highly restricted discharge. Since the threshold method always picks up the most sensitive fibre of all, this limitation is less serious than might have been imagined. In my latest experiments with the cat's retina, single fibres were obtained in 60% of the cases.

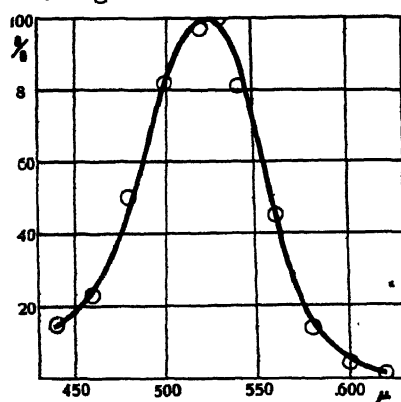


Figure 3. Spectral distribution of sensitivity of spike recorded in figure 2 (Granit, *Acta Physiol Scand.* 3, 1942.)

Success depends mainly on the success in making the microelectrode, and to some extent, of course, also on patience and practice. The animal must be quiet, for the slightest movement suffices to shift the microelectrode.

I have said that the fibre shown in figure 2 represented a simple type of

response. In the cat's eye, for instance, the fibres respond to an increase of stimulus intensity in a very complex manner. Most fibres discharge to both onset and cessation of illumination. But the relative amount of on- and off-discharges varies a great deal with stimulus intensity. Light may inhibit the on-discharge in a certain intensity range and inhibit the off-discharge in another. The complexity of the response, as soon as the threshold strength has been exceeded, is something stupendous. Hardly two fibres can be said to be exactly alike. An instance is given in figure 4. The figures to the right above each

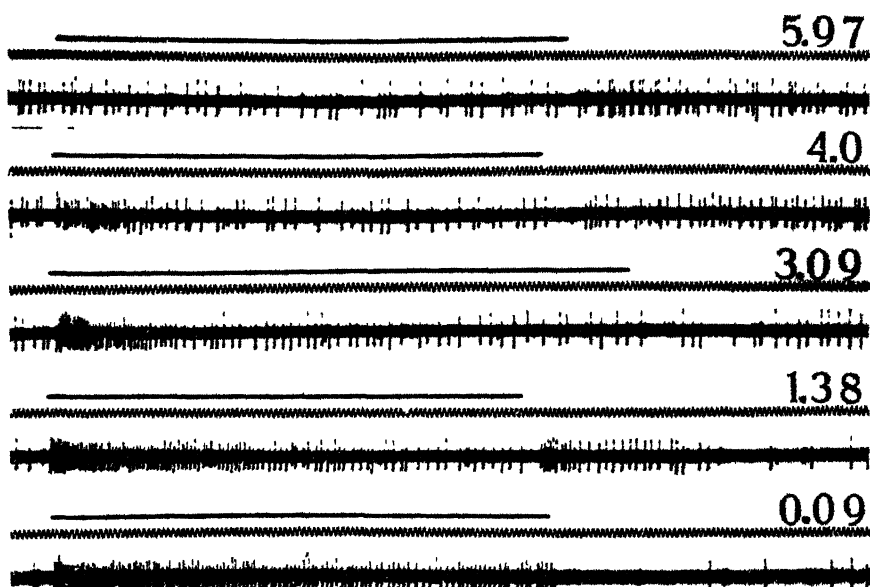


Figure 4. Cat Isolated spontaneously active element responding to progressively increasing intensities of white light. The numerals to the right show the extinction (densities) of the filter-wedge combination used in front of a light of 892 m.c. Marked as figure 2 (Granit, *J. Physiol.* 103, 1944)

record show the density of the neutral filter put into the beam of a white light of 892 m.c. In this case the fibre is spontaneously active all the time. The on-discharge increases with stimulus intensity (downwards), the off-effect increases first, is inhibited in record 3.09, increases a second time (record 1.38) and is ultimately completely inhibited at maximal intensity (record 0.09). These complex effects are due to interaction in the extremely complex retinal switchboard of neurons and synapses from which the optic nerve fibres take off. Such phenomena must play an important rôle for discrimination and contrast effects, but they do not concern us here. Figure 4 merely serves as a warning against premature and over-optimistic simplification when one is dealing with a wonderful microcosm of a nervous centre such as the vertebrate retina.

§ 3. THRESHOLD MEASUREMENTS FOR COLOUR-RECEPTION ANALYSIS

From the point of view of the analysis of colour reception we must begin with threshold measurements in the manner just described. For this immediate purpose it is immaterial how the discharge is modified by an increase of stimulus

intensity. We can always determine the energy necessary for a threshold response. This is the method that has been used in my analysis of the fundamental problem of colour reception, the problem raised with such clarity by the far-sighted genius of Thomas Young.

The results of measurements of the energy necessary for a threshold response at various wave-lengths could be very easily interpreted, were it not for the fact that in most of the laboratory animals both rods and cones converge towards the same fibre, as in figure 1. Now the rods contain the extremely light-sensitive substance *visual purple*, which has maximum absorption around $500\text{ m}\mu$ and a quantum yield, as determined by Dartnall, Goodeve and Lythgoe (1936), of 1, so that one quantum of light destroys one molecule of visual purple. Its absorption curve has been fairly well known since the work of Trendelenberg in 1904, and was first accurately determined with up-to-date technique of extraction and photo-chemical analysis by the late Dr. Lythgoe in this country in 1937, and almost simultaneously by several other workers. Schneider, Goodeve and Lythgoe showed in 1939 that the photosensitivity of visual purple agreed with the human scotopic luminosity curve, illustrating the distribution of brightness for a dark-adapted eye to a spectrum of low intensity. (For a full review of this work, see author's summary, Granit, 1946).

It is therefore not surprising that my method reproduces the same curve if the threshold energy in the spectrum is determined for fully dark-adapted eyes of animals containing rods charged with visual purple. Such measurements will be shown later. At the moment this point is taken up in order to emphasize that the visual purple distribution of spectral sensitivity is a serious complication in work with animals. Visual purple can be removed by light-adaptation, but it regenerates during the time the experiment is carried out in the dark, and so tends to raise the sensitivity in the green around $500\text{ m}\mu$. This, of course, is a consequence of the convergence of several receptors of mixed nature towards each optic nerve fibre as well as of the extreme sensitivity to light of visual purple. But when we use our own visual purple mechanism in the dark, the spectrum appears colourless. An increased sensitivity to green around $500\text{ m}\mu$ cannot, therefore, in the microelectrode experiment, be ascribed to a hypothetical "green response" of the *cones*. In order to be able to do so we must make certain that we have not been engaged in measuring merely the photosensitivity of visual purple. The experimental work becomes something in the nature of a fight against the absorption curve of visual purple, which tends to cover up the properties of other receptors belonging to the convergence unit which has been isolated by the microelectrode.

§4 THE DOMINATOR AND MODULATORS

In order to illustrate one of the main findings, I shall therefore first draw attention to a curve from the eye of a snake which is lacking visual purple. This animal has a pure cone retina. In figure 5 the curve interrupted by dots refers to the snake (1943 b), the other curve to the frog's eye (1941 c), which has been light-adapted so as to depress the activity of visual purple. The two curves are sufficiently similar to support a conclusion that after light-adaptation the remaining cones of the frog's eye have behaved like the cones of the snake. I shall refer

to this broad cone curve with maximum around 560 $m\mu$ as the *dominator*. Those familiar with the problems of vision will immediately notice that the dominator curve represents a distribution of sensitivity which is almost identical with the human daylight luminosity curve for cone vision, the so-called *photopic luminosity curve*. Provided that an animal's eye possesses a sufficient number of cones, this

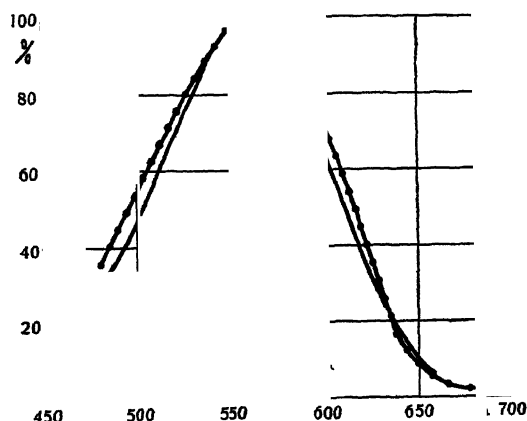


Figure 5. Distribution of sensitivity of dominator element in the retina of snake (line interrupted by dots) and frog (uninterrupted line) (Granit, *Nature, Lond.*, 151, 1943)

curve is always obtained, and represents the most common electrophysiological finding. It is lacking in the eyes of rats and guinea pigs (Granit, 1941 a, 1942). The guinea pig's eye is said to lack cones altogether, the rat's eye is said to contain about 1 % cones, apparently too few to be able to give the dominator.

In a sense the dominator can be described as a physiological unit response, but it is very probable that several receptors have combined to form this unit. We must remember that all optic nerve fibre units are convergence-units, as stated above.

When different optic nerve fibres are picked up, in the manner described, the whole experiment is very much dependent upon statistical chance and good luck. These factors favour the dominator, but other types of curve are also sometimes obtained. These are illustrated in figure 6. The curves are from eyes of different animals : rats (1941 a), guinea pigs (1942), frogs (1941 c), snakes (1943 b). First should be noted that they differ from the dominator curve in two important respects : they are much narrower and are spread over a large fraction of the visible spectrum.

Neglecting for the moment the narrow curve with maximum around 500 $m\mu$, which may be described as a narrow visual purple curve, the most striking fact demonstrated by figure 6 is the confinement of the narrow curves to three preferential regions of the spectrum. This is the more remarkable if one considers that they are from different animals. I have called these curves the *modulator curves*.

The most characteristic red modulator had its maximum at 600 $m\mu$. It was found in rats (dots), much to my surprise, since I took up the rat's rod eye in order to find out what happens to the visual purple distribution of sensitivity when the eye becomes light-adapted. The result was that I found the red modu-

lator remaining, when the green end of the spectrum, occupied by the visual purple absorption, had disappeared below the instrumental threshold of the Hilger-Tutton monochromator used at that time. After some time in the dark the narrow curve with maximum at $500\text{ m}\mu$ turned up. Later, during dark-adaptation, it expanded and assumed the shape of the typical visual purple absorption curve. I have since seen the same narrow curve in other rod eyes, but not in the cone eye of the snake. It is difficult to say whether it would play any part in human colour vision or not. Probably not, if central fixation is used. The rat's red modulator may be due to the 1% cones. Histologists, though not all, believe in cones in the rat's eye. There was no dominator in this eye. The red modulator was also seen in the eyes of frogs, in the pure cone eye of the snake, and was indicated as a hump on the green curve in some very rare cases in the guinea pig. In the frog I found some modulators with maximum at $580\text{ m}\mu$. These will be called yellow modulators.

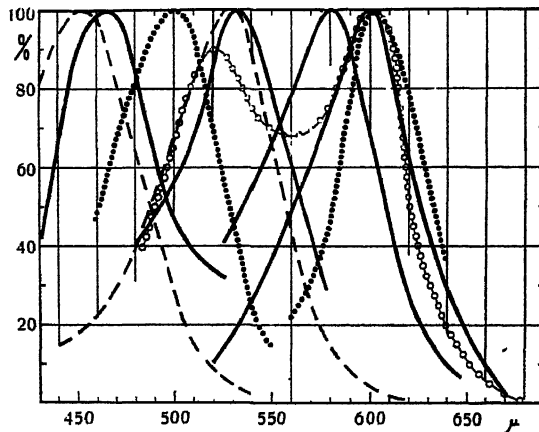


Figure 6 Distribution of sensitivity of modulator elements from eyes of rat (dots), guinea pig (broken line), frog (line in full) and snake (line interrupted by circles) In this figure and in figure 5 spectra of equal quantum intensity. The ordinates on either side of $560\text{ m}\mu$ drawn down to indicate dominator values (Granit, *Nature, Lond*, 151, 1943.)

Green modulators are found in all eyes between 520 and $540\text{ m}\mu$, even in the cone eye of the snake, but in this animal it proved impossible to separate them from the red modulator. Blue modulators around $460\text{ m}\mu$ were found in the frog's eye and in the pure rod eye of the guinea pig, but not in the cone eye of the snake.

The general biological implications of these findings are clear enough. Nature has not taken the trouble to invent new mechanisms of colour analysis for every new species. It has, as it were, decided upon the principles to be used and then proceeded to use them with what, from our point of view, we might call a greater or lesser degree of perfection. The eyes of the different animals are probably adapted to the life and habits of the species concerned. With regard to modulators, the boundaries between rods and cones do not seem to be very strict. It seems certain that dominators only are found in eyes with a large number of cones. It is possible and even probable that red modulators are cones in the histological sense, but modulators in the short wave-lengths need not be cones.

Blue modulators must be rods in the histological sense, and it is doubtful whether blue modulators can ever be anything else. The experiments suggest that, on the whole, more attention should be paid to photochemistry than to histology. Histological and photochemical definitions as to what is a cone or a rod may disagree. Unfortunately cone photochemistry is still at the stage when it is no exaggeration to state that the less said about it the better. It is, of course, very difficult to extract substances from receptors which to all appearance are practically colourless.

If I may state my own belief it is that the cone substances are modifications due to changes in the bonds linking the visual purple chromophore to its protein body. There is some evidence for this view which cannot be discussed in this connection (see Granit, 1941 b). Some day those working on the photochemistry of visual purple will provide us with important clues to the solution of this riddle. Lythgoe made a beginning when, shortly before his death, he found that visual purple, regenerating from its photoproduct "transient orange", had its absorption curve shifted towards the long wave-lengths. This work was never published.

§5 NATURE OF THE DOMINATOR

The microelectrode experiments now described did not suffice to create that feeling of satisfaction with which the completion of a work properly should end. The element of statistical chance involved made them not only extremely tedious but also unsatisfactory; and what was the meaning of the dominator, which was the most common finding in eyes containing a great number of cones? Could it be regarded as being formed by grouped modulators? Those questions

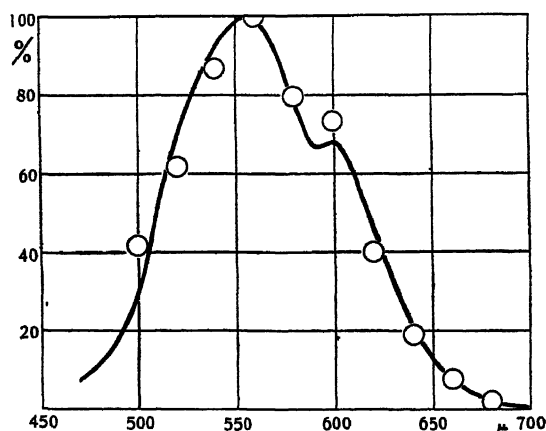


Figure 7. Luminosity curve determined by Wright (*Nature, Lond*, 151, 1943) with a small fovea patch of low brightness. The circles illustrate the cat dominator from Granit (*Acta Physiol. Scand.* 5, 1943 a.) Equal-energy spectrum.

became more pressing when it was found that in some cases the dominator of the snake eye had a hump around 600 mμ and that all dominators in the cat's eye had this same hump. Despite this it proved impossible to obtain red modulators in the cat's eye, although this eye is more like the human peripheral retina than any other eye used in my work.

At that time Dr. Wright published in *Nature* (1943) a brief note reporting the results of an experiment in which he had tried to imitate the microelectrode

method by measuring the luminosity curve of a very small patch of light. He also found a hump on the foveal luminosity curve in the red around 600 $m\mu$. Figure 7 shows Wright's curve. The circles refer to my own measurements of the cat's dominator. This parallelism between two independent measurements with very different techniques suggested that the dominator might be a composite curve, besides suggesting that the human and the feline mechanism of colour reception cannot be fundamentally different. It therefore became imperative to develop a method by means of which it would be possible to split the dominator, if it could be split. Considering that about 36% of the isolated fibres in the light-adapted cat's eye gave the dominator, a successful splitting of the dominator would minimize the statistical element of chance involved in the process of hunting for simpler units of colour reception. For the new experiments I used Dr. Wright's well-known colorimeter (1934), drawn and constructed for this work by Mr. G. C. Newton.

The principle of the experiment is illustrated in figure 8. The animal is

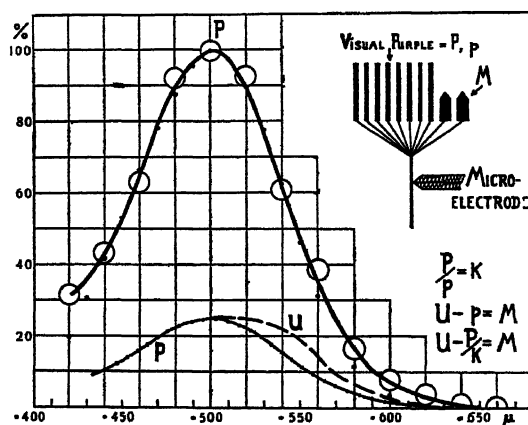


Figure 8 Dark adapted cat. Large circles=average distribution of sensitivity of isolated fibres in the optic nerve. Small black circles=Lythgoe's corrected curve for visual purple absorption from *J Physiol.* 89, 1937. Spectrum of equal quantum intensity. As to significance of diagram and symbols, see text. (Granit, *J Neurophysiol.* 8, 1945.)

fully dark-adapted so that the highly sensitive visual purple completely dominates and gives the curve P . This is based on some 1320 readings, averaged into 179 values, again averaged to give the large circles around the curve. The small black points are Lythgoe's corrected curve for visual purple absorption, which is a little too low in the violet, as shown by Schneider, Goodeve and Lythgoe (1939). A slight effect of the dominator makes my curve a little too high in the yellow-red region.

If the eye be adapted to red, blue or green light the curve P is reduced to p , by proportionate ordinates at all wave-lengths, since P represents the homogeneous substance visual purple. Hence

$$P/p = k. \quad \dots\dots(1)$$

But if there are any other colour-sensitive substances (M), *preformed* or produced by visual purple, curve P does not drop to p but to some other curve, U

(figure 8). The modulators M are given by the difference between U and p , as evident from the diagram, or

$$M = U - p. \quad \dots\dots(2)$$

In this case p is unknown, but the equation can be solved by giving it the form

$$M = U - P/k \quad \dots\dots(3)$$

P is obtained *before*, U *after*, selective adaptation, and k can be obtained from (1) by finding the spectral region in which selective adaptation has caused the largest drop of sensitivity, since in this region there was no other substance than visual purple left to resist light-adaptation. This region will generally be found in the place where the letter p is inserted in the diagram (figure 8). All quantities of (3) are now known, and the argument can be tested by experiment.

Equation (3) was solved in 34 experiments on the basis of some 4000 observations collected in 601 points on U -curves. Some 60% of the U -series referred to isolated fibres, the rest to restricted activity. In 29% the equation came out zero ($U=p$), and hence the microelectrode had struck a unit with pure visual purple receptors. In the rest of the series complex curves were obtained.

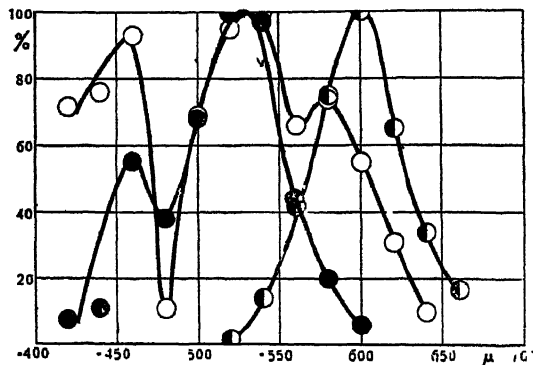


Figure 9 Average effects of selective adaptation of cat's eye, as described in text. ● = red adaptation, ○ = green adaptation, ● = blue adaptation. Spectrum of equal energy (Granit, *J. Neurophysiol.* 8, 1945)

Coloured adaptation was used because it was hoped that, for instance, red adaptation would suppress red-sensitive elements and give elements sensitive to other colours a chance to appear. The eye was left in coloured light for some time, this light then instantaneously interrupted, and a test light from the spectrum flashed in, 3 sec. after interruption of the coloured adaptation. This time was chosen in order to give the off-effect time to disappear or to diminish in frequency, so as to make it possible to hear whether the test light caused a fresh discharge or not. When the test had been carried out, the adapting light was again switched on for a while until the eye was ready for a new test with some other spectral wave-length. For adaptation the Ilford spectral filters red, green and blue were used in the main series. A minor number of experiments were carried out with the yellow and violet filters.

For a general survey of the M -curves from equation (3), figure 9 should be studied. This does not show any individual modulators, but merely the general

effect of adaptation to red, green and blue light as gross averages of the *M*-curves obtained. The maximum for each individual experiment has always been given the value 100, so that the curves give an idea of the chance for a given point to reach a certain magnitude.

It is seen that blue adaptation gave the most uniform results, so that only curves with maximum in the red were obtained. Red adaptation actually suppressed red modulators, but green and blue modulators made themselves felt. The adaptation to green had a still less selective character. There were humps in both the red, green and blue regions of the spectrum.

Let us now present the same results in a manner which is easier to follow. The individual modulators from the three preferential regions have been picked out and averaged, independently of the kind of coloured adaptation by which they have been obtained. The results are shown in figure 10, in which, also, the dispersion is indicated (outer contours). The narrow red modulators were of two types with maxima respectively at 600 $m\mu$ (red) and 580 $m\mu$ (yellow), the former type being the more common. Most green modulators overlapped and had

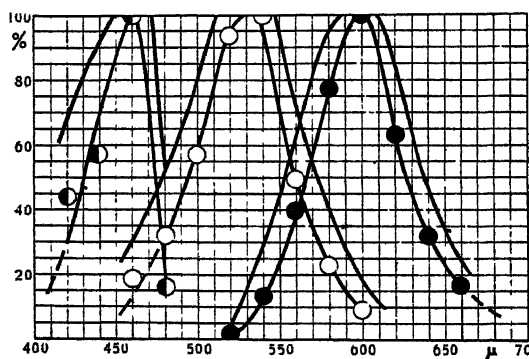


Figure 10. Averaged individual modulators (cat) as obtained by selective adaptation. ●=red modulators, ○=green modulators, ◐=blue modulators. Outer contours indicate dispersion, see text. Spectrum of equal energy (Granit, *J Neurophysiol* 8, 1945)

maxima at 540 $m\mu$, some at 520 $m\mu$, and two of them were of the type previously described as narrow visual purple curves. These may be lacking in pure cone eyes. These modulators were the ones seen before in different types of retinae, analysed by the earlier "chance" method. The blue region could be better analysed with Wright's colorimeter. Most blue modulators had their maxima at 460 $m\mu$ (blue), one at 440 $m\mu$ (violet). They were very narrow bands. Errors of measurement increase from the red to the blue end. On account of the steepness of the modulator curves the maxima are fairly well definable.

The average modulators of figure 10 are too narrow to add up to a cat dominator. It is clear that impulses also must be delivered up the optic nerve by modulators with "legs" outside the averages. The dispersion gives some idea of the extension of the outer margin of each of the three groups of modulators. In figure 11 the extreme values obtained in these experiments are given. Together with the curves of figure 10, showing the dispersion, they should give some idea of the limits permitted in synthesizing the dominator.

In figure 12 is shown the synthesis of the cat's dominator (Granit, 1943 c),

from figure 7. The thick line, drawn through the readings, is the sum of the red and green curves, R and G. In the light of these experiments it seems permissible to maintain that the dominator is a composite curve consisting of modulators. Nevertheless the dominator must be regarded as a *biological unit*. It cannot be neglected in theories of colour vision even though the modulators are

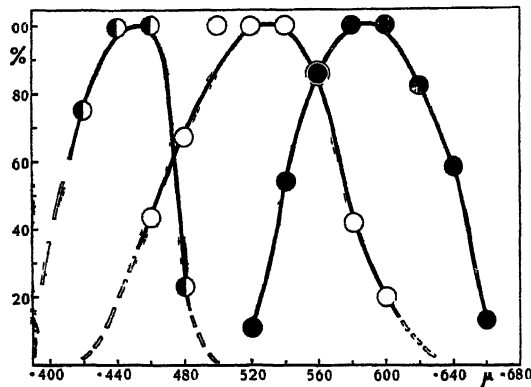


Figure 11. Extreme values obtained for cat modulators by selective adaptation, as described in text. Equal-energy spectrum. (Granit, *J. Neurophysiol.* 8, 1945.)

the *ultimate physiological units*, the units of first order. It should be emphasized that modulators in the pure state were also obtained without coloured adaptation in some animals.

Why have modulators never been found in the human eye with sensory methods? The main reason would seem to be that the analytical unit-field

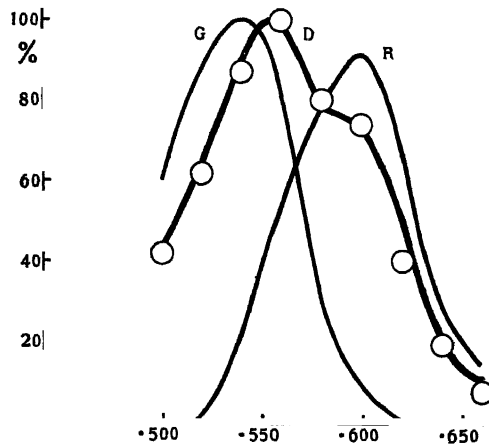


Figure 12. Circles = averages of four photopic dominators of cat. Curve D in heavy lines = synthesis of this dominator by adding modulator areas G and R and plotting their sum in per cent of maximum Equal-energy spectrum (Granit, *J. Neurophysiol.* 8, 1945)

contains too many elements, perhaps up to a hundred thousand nerve fibres. I should also like to draw attention to the fact that the microelectrode technique is a threshold measurement. Wright's and Walters and Wright's (1943) promising results suggest that experiments with small areas and low intensities, perhaps aided by selective adaptation, might open up new possibilities. Some-

thing might also be gained by studying the peripheral field of vision (see Walters and Wright, 1943).

In figure 13 I have synthesized the human photopic luminosity curve on the basis of the results obtained with the cat's optic nerve. The daylight spectrum or cone spectrum of the human eye agrees very well with the dominator (D) of the electrophysiological experiment. The three preferential regions within which modulators are found are given by the R, G and B curves, forming the three fundamental sensation curves of the Young-Helmholtz trichromatic theory.

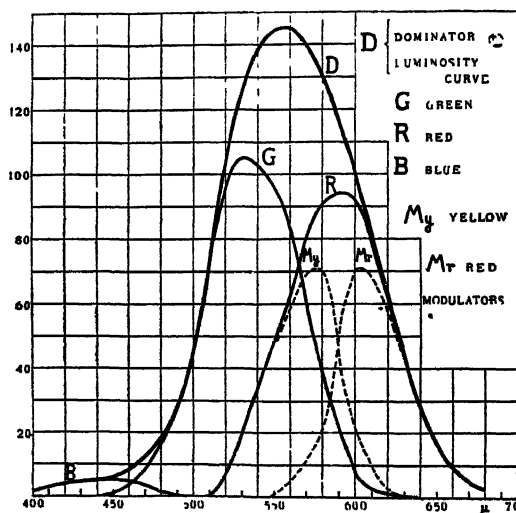


Figure 13. Synthesis of human photopic luminosity curve (D), as determined by Coblentz and Emerson (*Bull. Bur. Stand., Wash.*, no. 303, 1917), on the basis of three fundamental sensation curves, R, G and B. The R-curve indicated to be the sum of two modulators M_y and M_r . Equal-energy spectrum. (Granit, *J. Neurophysiol.* 8, 1945.)

Their sum is the dominator. For the R-curve I have indicated the two modulators M_r and M_y . It would, no doubt, be possible to combine the modulators in a slightly different manner, for instance an R-curve with a hump at 600 $m\mu$. But the last word on this question must remain with those who are experts on human colour psychophysics. It should be remembered that threshold measurements, such as are used in electrophysiological measurements, record the narrowest curve that can be obtained. It seems possible that the greater the level of intensity, the greater the width of the modulator curve. The retina, after all, is a complex nervous centre with mechanisms of facilitation.

§ 6. COLOUR VISION

It should be emphasized that the electrophysiological experiments have not provided any evidence for the existence of three fundamental response curves. From the physiological point of view, colour vision must be understood in terms of modulators and dominators. The three fundamental response curves must be regarded as approximations depicting the areas covered by the modulators in the three preferential regions. But it is interesting to note that my schematic response curves agree fairly well with those obtained by Walters (1942) in experi-

ments developing Wright's (1934) method of selective adaptation. Walters' curves are shown in figure 14.

Recent work by Pitt (1944) on the three response curves has not yet become available to me, but from a summary by Stiles (1944) I infer that Walters' results have been confirmed by Pitt by a different method. Pitt's curves extend further down into the blue part of the spectrum.

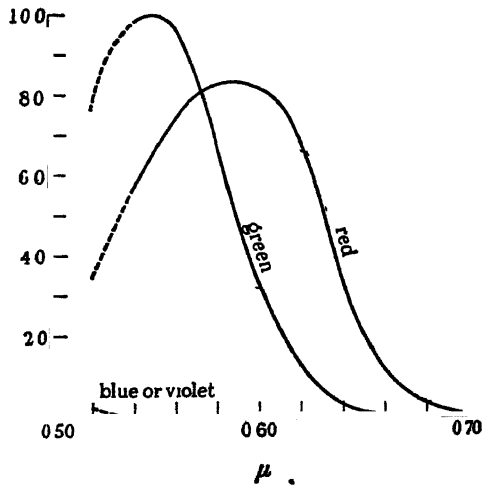


Figure 14 Fundamental sensation curves determined by Walters (*Proc Roy. Soc B*, 131, 1942)

The fineness of hue discrimination would seem to be well explained by the narrow modulator curves. Figure 15 is taken from the work of Wright and Pitt (1934). The ordinates show the shift in wave-length, necessary for discrimination of hue at adjacent wave-lengths, so that minima in the curve signify maxima of discrimination. In the region between 580 and 600 $m\mu$, where the red and the yellow modulators intersect, one can distinguish wave-lengths separated by as little as 1 $m\mu$. From figure 13 it is clear that precisely in this region a slight shift of wave-length suffices to redistribute the nerve fibre signals between

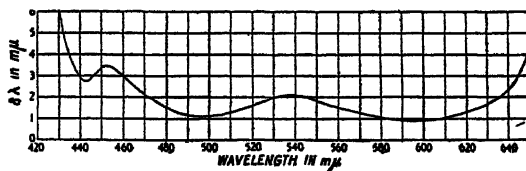


Figure 15. Human hue discrimination according to Wright and Pitt (*Proc Phys Soc* 46, 1934)

the red and the yellow modulator. The second optimum between 480 and 510 $m\mu$ is found in a place where the blue and red modulators meet and intersect with the green ones. Rather remarkable is the third (less marked) optimum between 450 and 440 $m\mu$, which suggests, as did my experiments, that two blue modulators overlap in this region, the blue proper with maximum at 460 and the violet one with maximum at 440 $m\mu$.

§7 THE NATURE OF WHITE AND COLOUR

The existence of the dominator as a physiological unit makes it necessary to consider its sensory equivalent. It is well known that our sensations of light can be divided into two main categories, brightness or luminosity and colour or hue. The luminosity curves of the human eye have several times been referred to. It is well known that corresponding to our two sense-organs in the retina, the rods and the cones, there are two luminosity curves, one for dark-adapted rods charged with visual purple, another for the light-adapted eye and cone vision. These two standard curves are shown in figure 16. The daylight luminosity curve to the right we have identified electrophysiologically as the photopic dominator, the scotopic luminosity curve to the left as a replica of the visual

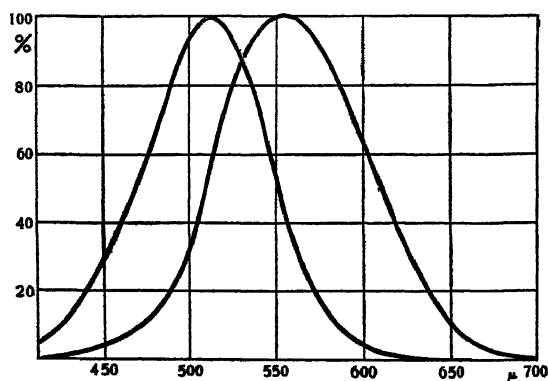


Figure 16 The standard photopic and scotopic luminosity curves of the human eye

purple absorption curve minus losses of light of short wave-length in the ocular media. If the latter curve is corrected for this and plotted on a quantum basis, it will actually be identical with the curve for visual purple absorption.

Figure 16 is introduced here merely in order to show that a shift of the maximum need not as such cause an impression of colour. Both curves are luminosity curves illustrating the distribution of brightness. From the physiological point of view it seems reasonable to suggest that optic nerve fibres representing wide spectral areas, such as those of figure 16, cannot cause any other impression than that of brightness or whiteness. For discrimination of wave-length, nature has developed a very different type of mechanism, consisting of narrow modulators in three preferential regions.

According to this view the dominator represents brightness or whiteness. To put it differently, the fundamental visual equivalent to impulse frequency is a neutral white. Colour, if I may say so, is a kind of "local sign", carried by the modulators, which are certain receptors or groups of receptors connected to certain fibres so that from the retina up to the brain the same spatial pattern of organization is maintained. Inasmuch as the modulators contribute to impulse frequency they can also contribute to brightness.

Galambos and Davis (1942) have recently applied the microelectrode technique also to the acoustic nerve. There too it was found that the different acoustic frequency bands were represented by different fibres, rather narrow bands at that. In the ear too the perception of quality or pitch is something

related to space, just as in the eye. Loudness, however, is almost certainly a function of total impulse quantity, just as brightness in the eye. Colour and pitch are determined by some kind of local sign.

What makes the eye particularly interesting is that the spectral distribution of brightness is also found to be represented by individual optic nerve fibres. In my opinion this must be regarded as evidence for the view that white has the character of a separate sensation. It has long been known that brightness and colour have maintained a certain amount of independence. This is particularly obvious in defective colour vision. Thus the deuteranope is a red-green blind observer with a normal distribution of luminosity. I do not see how this can be explained by the trichromatic theory, according to which white must be regarded as the sum of the red and green fundamental response curves with a minor contribution from the blue curve. If the red and green curves are lacking, one must expect a grossly abnormal distribution of luminosity instead of the normal one found. This difficulty does not exist for the dominator-modulator concept.

The various forms of colour blindness I regard as defects in the receptors or elsewhere in the paths representing the modulators. But such defects would have to be of an extreme degree in order to be so complete as to remove the dominators, of which, apparently, there is a very great number, considering how easily they can be detected.

There are, too, several forms of acquired colour blindness. They are characterized by loss of perception of certain hues but not of brightness. In certain forms of colour blindness, such as protanopy, there are minor shifts in the luminosity curve. These can be explained by a minor redistribution of the aggregate modulators forming the dominator.

The distribution of saturation in the spectrum is very characteristic. The ends are saturated and at the same time dark, extreme red and extreme blue being examples of such dark and saturated colours. On the dominator-modulator concept this is explained by the fact that at the ends of the spectrum the dominator values are very low, so that the modulators in these regions are responsible for the greater part of the sensory experience. Again, in the yellow region the large contribution from the dominator makes the spectrum appear relatively white and the colour unsaturated.

§ 8. CONCLUSION

Let us now return to Thomas Young, in whose honour this lecture is being delivered. It is a characteristic of those whom we revere as classics of science that their ideas have been formulated with a curious insight into how nature might be expected to behave and so have preserved lasting creative power. What greater tribute could one scientist pay to another's memory than to perform an experiment suggested by his ideas, 140 years after they have been formulated, and come to the conclusion that these ideas were fundamentally correct? The mechanism of colour reception is organized by the peripheral visual apparatus, the number of colour-sensitive elements is relatively limited, and these elements represent widely different regions of the visible spectrum. Those were Young's three fundamental assumptions. He was right even in assuming three main types of colour-receiving apparatus. These are the three preferential regions

within which modulators are found. The electrophysiological work may, indeed, be said to have confirmed the view he gave of the framework of a mechanism of colour reception. Its finished picture looks somewhat different, but the old framework was solid enough and shines through.

It has been a great pleasure for me to have had this opportunity of speaking to this distinguished audience in honour of Thomas Young. His original theme was the same as mine, in the sense that he described a peripheral mechanism of wave-length reception in terms of the properties of nerve fibres. Colour vision is a much wider subject, and there are several in this Society who know much more about it than I do. My theoretical attempts to translate the electrophysiological results into the language of colour psychophysics are legitimate, and I believe them to be correct, but further experience may nevertheless necessitate modifications. I can only hope that I shall not have to make these experiences myself, but that somebody else will try his hand at the optic nerve. I also feel just now that it would be interesting to see for a while what photochemistry and colour psychophysics could do for this field before any further labour is invested in electrophysiological work. For the invitation to present the results of this work to the Physical Society I wish to express my sincerest thanks.

REFERENCES

- ADRIAN, E. D., 1932 *Mechanism of Nervous Action* (Oxford: University Press).
 ADRIAN, E. D. and MATTHEWS, R., 1927 *J. Physiol.* **63**, 378.
 COBLENTZ, W. W. and EMERSON, W. B., 1917 *Bull. Bur. Stand., Wash.*, no. 303.
 DARTNALL, H. J. A., GOODEVE, C. F. and LYTHERG, R. J., 1936 *Proc. Roy. Soc. A*, **156**, 158.
 GALAMBOS, R. and DAVIS, H., 1942. *J. Neurophysiol.* **6**, 39.
 GRANIT, R., 1941 a *Acta Physiol. Scand.* **2**, 93.
 GRANIT, R., 1941 b *Acta Physiol. Scand.* **2**, 334.
 GRANIT, R., 1941 c *Acta Physiol. Scand.* **3**, 137.
 GRANIT, R., 1942 *Acta Physiol. Scand.* **3**, 318.
 GRANIT, R., 1943 a *Nature, Lond.*, **151**, 11.
 GRANIT, R., 1943 b *Acta Physiol. Scand.* **5**, 108.
 GRANIT, R., 1943 c *Acta Physiol. Scand.* **5**, 219.
 GRANIT, R., 1944 *J. Physiol.* **103**, 103.
 GRANIT, R., 1945. *J. Neurophysiol.* **8**.
 GRANIT, R., 1946. *Sensory Mechanisms of the Retina* (Oxford: University Press. In course of publication).
 HARTLINE, H. K., 1938 *Amer. J. Physiol.* **121**, 400.
 HARTLINE, H. K., 1940 *J. Opt. Soc. Amer.* **30**, 239.
 LYTHERG, R. J., 1937. *J. Physiol.* **89**, 331.
 SCHNEIDER, E. E., GOODEVE, C. F. and LYTHERG, R. J., 1939. *Proc. Roy. Soc. A*, **170**, 102.
 STILES, W. S., 1944 *Proc. Phys. Soc.* **56**, 329.
 TRENDLENBERG, W., 1904 *Z. Psychol. Physiol. Sinnesorg.* **37**, 1.
 WALTERS, H. V., 1942 *Proc. Roy. Soc. B*, **131**, 27.
 WALTERS, H. V. and WRIGHT, W. D., 1943. *Proc. Roy. Soc. B*, **131**, 340.
 WRIGHT, W. D., 1934 *Proc. Roy. Soc. B*, **115**, 49.
 WRIGHT, W. D., 1943 *Nature, Lond.*, **151**, 726.
 WRIGHT, W. D. and PITT, F. H. G., 1934. *Proc. Phys. Soc.* **46**, 459.
 YOUNG, T., 1855. *Miscellaneous Works of the late Thomas Young*, ed. by George Peacock, vol. 1 (London).

THE GEOPHYSICAL ASPECT OF COSMIC RAYS

BY ARTURO DUPERIER

London : formerly University of Madrid

Twenty-ninth Guthrie Lecture, delivered 5 July 1945

§1. INTRODUCTION

IN my student days I intended, like your founder, to become a chemist. Later, however, physics attracted me in consequence of my interest in thermodynamics, whereby I became acquainted with Professor Guthrie's important discovery of cryohydrates. I could hardly have imagined that the future would hold for me the honour of delivering one of the lectures instituted by this distinguished Society in his memory. This invitation by your Officers and Council is another proof of the kindness shown to me by this country, whose generosity has enabled me to carry on my scientific work.

On the suggestion of Professor P. M. S. Blackett, I undertook in 1939 the study of the variations of cosmic-ray intensity with time. All work carried out in recent years has shown increasingly the close relation of this study with that of terrestrial magnetism and with the physics of the atmosphere. In presenting to you some of the results I have obtained, I shall try to give you an outline of the present state of our knowledge of this geophysical aspect of cosmic rays.

§2 THE TEMPERATURE EFFECT OF COSMIC RAYS

In the last decade or so cosmic rays have been recorded in various parts of the world. The early records, obtained by using ionization chambers, showed an annual variation of the intensity of cosmic rays inverse to that of the temperature near the ground.

It is well known that Blackett (1938) attempted to relate this temperature effect with the instability of the mesons, which form the main part of the penetrating component, and that in this way he explained the decrease of cosmic-ray intensity in warm weather as being due to the greater distance which the mesons have to travel to reach sea-level owing to the greater height of the pressure-level at which they originate. As Blackett pointed out, to test this theory the observed intensity changes would have to be correlated with the mean temperature of the free atmosphere rather than with the temperature near the ground, as had been done until then.

Duperier (1941), using various published cosmic-ray records with the upper-air temperature data for Europe compiled by Wagner and those for the United States compiled by Lennahan, found that the seasonal cosmic-ray variations are related more closely to the average spatial temperature of the atmosphere up to 16 km. than to the temperature near the ground. In particular, it was shown that the lag in the warming of the atmosphere in spring is paralleled by a lag in the diminution of intensity of cosmic rays. But though this was the only way of treating the data then available, it was clear that such a correlation had but

slight physical significance, particularly as the observations of atmospheric temperature referred to different intervals of space and time from those of cosmic-rays.

At about the same time, Beardsley (1940) correlated the cosmic-ray intensity at Cheltenham (U.S.A.) and the upper-air temperatures as obtained by radiosonde balloons released during the period covered by the observations. He found that up to 10 km. the correlations were not appreciably different from those based on ground temperature, but that above this altitude the correlation becomes less close as the height increases, and even changes sign. This would mean that the temperature averaged over a height of 15 km. would be less closely correlated with cosmic-ray intensity.

Later, Hess and Benedetto (1941) and Benedetto, Altmann and Hess (1942) made a similar investigation using the data obtained with a cosmic-ray dual telescope and upper-air temperatures which were observed during the same period. They first correlated cosmic-ray intensity at ground level with temperature values at different levels from the surface up to 12 km., and found that the correlation decreased as the height increased. Because of this result, they were led to a new method of taking mean temperatures. Instead of the ordinary method of integrating the temperature $v.$ height curve and dividing by height, they adopted the method of integrating the temperature $v.$ pressure curve and dividing by pressure. To the temperature thus defined they gave the name of *mass temperature*. Then they correlated the mean daily cosmic-ray intensity with the mean mass temperature for increasing fractions of the atmosphere, and found that this correlation increases, though only very slightly, as the fraction of the atmosphere becomes greater. from $r = -0.71$ for 0.2 to $r = -0.75$ for 0.8 of the atmosphere.

Clearly, from these results it would not be possible to say whether the so-called temperature effect of cosmic rays is controlled by the mean temperature of the atmosphere up to a certain level rather than by the temperature near the ground, however the temperature be defined.

But the lack of success in proving the predominant influence of the atmospheric temperature was perhaps not surprising. As a result of the instability of the meson, it is the change in height of the pressure-level at which mesons are formed which gives rise to the fluctuations of the penetrating cosmic rays at the ground, and, as the hypsometric formula shows, the height of any pressure-level depends not only on the temperature, but also on the vertical distribution of temperatures in the air underneath and on the pressure at sea-level. When, in order to find the correlation with temperature, the ground-level meson intensities are first corrected for pressure by using, as is customary, the so-called barometric coefficient, the effect of the pressure change at sea-level on the height is removed, but there still remains the effect of distribution of temperatures to be added to the effect produced by change of mean temperature. Now, since the upper-air observations show that the differences in temperature distribution from day to day are generally far from being negligible, it must be admitted that the simple correlation of daily cosmic-ray changes with mean temperatures cannot lead to convincing results.

On account of this, I have (1944) adopted another method for the analysis of the cosmic-ray records which have been obtained in London in the last few years,

by making use of a battery of Geiger-Muller counters registering triple coincidences.

Cosmic rays travelling through the atmosphere lose energy in several ways, the amount depending on the mass of air traversed. In addition, if we accept the principle of the instability of the meson, the penetrating rays are also subject to decay. We may thus establish that the variation of the cosmic-ray intensity at ground level is a function of, firstly, the variation of the mass of air (represented by the barograph reading), and, secondly, the change in height of the pressure-level at which mesons are generated, assuming that there is only one meson-generating layer. Now should this be correct, there must clearly be a correlation between the variations in intensity of cosmic rays and those in height of this pressure-level. But as we do not know where mesons originate we have to put to the test a few pressure-levels for which meteorological information is available.

For the computation the hourly numbers of cosmic rays and the hourly barograph readings were averaged in groups of 24 hours. Of these daily mean data, those corresponding to periods of great geomagnetic disturbance were discarded as well as those corresponding to days on which the sounding balloon at the meteorological station failed to reach an altitude of 16 km., so that the correlations of cosmic rays with the heights of different pressure-levels up to this limit will be entirely comparable.

The following table gives the values of the partial correlations for the pressure-levels which have been chosen.

Table 1

Pressure-level	Partial correlation coefficient
75 mm Hg (16.1 km)	-0.67
113 " (13.5 ")	-0.54
188 " (10.3 ")	-0.32
375 " (5.5 ")	-0.30

The gradual increase of the correlation with height shown by the table proves that it is not the temperature near the ground, or at least that this is not the only factor, which is responsible for the temperature effect of cosmic rays, since the influence of this temperature on the height of a certain pressure-level decreases with the mean height of this pressure-level. On the other hand, the significance of the value -0.67 for the correlation between cosmic rays at a constant pressure and the heights of the 75-mm. pressure-level can be taken as confirming the view that a part of the variation of the intensity of the cosmic radiation at ground-level may be explained by spontaneous disintegration of mesons in the atmosphere.

In order to check these results by means of other experiments, we may assume that mesons originate at this pressure and write the regression equation

$$C - C_m = \mu(B - B_m) + \mu'(H - H_m),$$

where C , B and H are the number of cosmic rays, barograph readings and heights respectively, and subscript m refers to mean values. Clearly μ represents the

true absorption coefficient in air, and μ' the mean rate of decay of mesons. By solving this equation in the usual manner, we find (Duperier, 1944) for the first coefficient $\mu = 2.28\%$ per cm. Hg, a value which is practically the same as the value that can be deduced from the measurements by Ehmert (1937) of the absorption curve in water.

For the mean rate of decay of mesons we obtain $\mu' = 5.4\%$ per km. This coefficient is the reciprocal of the so-called *mean range* of mesons before disintegration, and is related to the lifetime of mesons when at rest, τ_0 , by the very well known equation $\mu' = \frac{1}{L} = \frac{cM}{E\tau_0}$, where M is the rest mass and E the energy

of the mesons. If we assume that the mesons with which we are concerned have a mean energy of 3×10^8 ev., and take M equal to 200 times the mass of an electron, we have $\tau_0 = 2.1 \times 10^{-8}$ sec., which is of the same order of magnitude as the values found in other modern experiments, which generally vary between 2 and 3×10^{-8} sec.

These results suggest the possibility of applying the cosmic-ray records at the ground to foretell the daily mean height of the 75-mm. pressure-level and thence, by means of the hypsometric formula, the mean temperature of the air up to about 16 km. The calculation of the height would be made as follows.

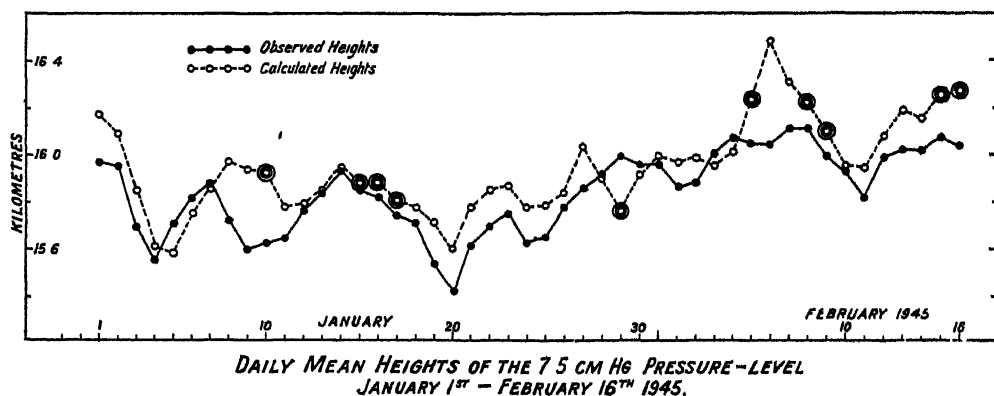


Figure 1.

We should first reduce the daily average cosmic-ray intensity to a constant pressure by making use of the *true* absorption coefficient. Then the percentage departures of these averages from the mean value divided by the mean rate of decay of the mesons would give the fluctuations in height of the pressure-level.

Figure 1 enables us to compare the heights as calculated in this manner with the actual heights as observed with sounding balloons for the period 1 January to 16 February 1945. The computation of the above coefficients had been made long before. The open double circles indicate the days on which a geomagnetic disturbance was recorded at Abinger. As the diagram shows, the correspondence between the two curves is, on the whole, striking, but on days of magnetic activity, and on the immediately preceding or following days, the discrepancies are both considerable and irregular. We should therefore run the risk of making a grave error in trying to foretell the mean temperature of the atmosphere overhead from the variations of cosmic-ray intensity at the ground.

However, as we shall see later, there is now some evidence that these cosmic-ray disturbances at periods of magnetic activity are of a world-wide character, having probably the same value at points on the same geomagnetic latitude. If this is so, the difference between the cosmic-ray intensities at two stations in a narrow zone of latitude, by eliminating the superimposed disturbance, would give a measure of the difference in temperature of the atmospheres over the two stations. Clearly, if the temperature of the atmosphere at a certain place was known by observation, it would be possible to determine the temperature of the atmosphere at all points in a narrow zone of latitude within which cosmic-ray records were being made. For this, it would, of course, be necessary for all the recorders to be sufficiently stable to give consistent results.

§ 3. DIURNAL VARIATION

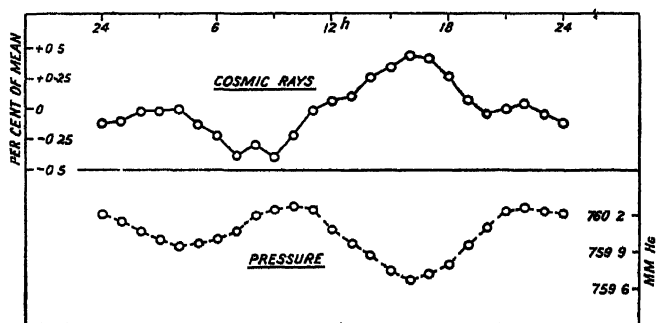
Besides the annual variation in cosmic rays which results from the effect of the thermal expansion and contraction of the atmosphere on the number of mesons reaching ground level, records show an appreciable change in cosmic radiation with the time of day. This solar diurnal variation in intensity has been measured by many observers, generally by using lead-shielded ionization chambers, and the results agreed in giving a maximum intensity around noon with an amplitude of about 0.2%. It has been also found, in particular by Thompson (1933) from the extensive measurements made by Compton and Turner during a year on the Pacific Ocean between latitudes 50° N. and 40° S, that both the amplitude and phase are independent of the latitude.

The fact that the maximum intensity has always been found to occur at about midday proved, without doubt, that the temperature effect cannot be responsible. The change in height of the meson-producing layer due to the daily variation in temperature, if appreciable, would give rise to exactly the opposite fluctuation. Hess (1936) pointed out the possibility of the existence of a component of cosmic radiation coming directly from the sun. Later Vallarta and Godart (1939) gave a theory based on the effect on primary cosmic rays of a magnetic field at the surface of the sun of the required magnitude to prevent cosmic-ray particles of less than 2×10^9 ev energy from reaching the earth. By assuming that the positively charged particles of the incoming radiation were in greater number than the negative ones, these workers found a diurnal variation at latitudes higher than 40° with the maximum occurring during the early afternoon, in agreement with the observed curve. If the ratio of positive to negative particles is taken as $\frac{4}{3}$, the amplitude also agrees with the measurements. For latitudes lower than 40°, Vallarta and Godart found that the effect of the sun's magnetic field became negligible. More recently, Jánossy and Lockett (1941), after analysing the effect on primary cosmic rays of different hypothetical heliomagnetic fields, came to a similar conclusion—that the diurnal variation at 45° latitude can be accounted for by assuming a magnetic field at the surface of the sun. They found also that the amplitude of the diurnal variation must be least in autumn and spring and greatest in summer and winter.

When the harmonic analysis of the observed curves is made, however, the results of different investigators are not always in agreement. It is found that there is a significant 24-hour wave with an amplitude of about 0.2%, which is approximately the same in all cases, though the time of the maximum can vary

by as much as eight hours from one station to another. In the same way, the possible existence of a 12-hour wave with an amplitude of about 30% that of the first harmonic is not excluded by some investigators. Forbush (1937) thought that the variations of the 24-hour wave might be due to variations in gamma radiation with locality. According to him, Wait has obtained results indicating the possibility of a daily change of ionization in the air due to gamma radiation, with a maximum at about noon. In passing through the 12-cm. lead shield the gamma rays would produce ionization inside the chamber which would give rise to an additional wave having an amplitude one-third that observed in cosmic rays.

I should like now to present the results, not yet published, that I have obtained at the Imperial College concerning the diurnal variation of cosmic-ray intensity by using the data for the period May 1941 to April 1944. The data for days of great geomagnetic activity have been discarded. Figure 2 shows the average hourly numbers of threefold coincidences, or cosmic-ray particles, as percentages



DIURNAL VARIATIONS OF COSMIC-RAY INTENSITY AND ATMOSPHERIC PRESSURE AT LONDON

Figure 2

of the mean. As the total number of coincidences recorded for each hour of the days within three years amounts to about 24 millions, the statistical fluctuation has the very small value of 0.02%. The box which contains the set of Geiger-Muller counters is kept at constant temperature by thermostatic control. The electric clock of the recorder has been compared with a chronometer at intervals of a few days over a period of several months, and no systematic difference between day-time and night-time has been found. The variation shown by the diagram can therefore be considered as physically real. Figure 2 also illustrates the average daily variation of pressure in mm. Hg obtained from my hourly readings during the same days of the three-year period.

The harmonic analysis of both curves gives, for the 24-hour wave,

$$\text{Cosmic rays } 0.25 \cos \frac{360^\circ}{24} (t - 17.3) \% , \quad \text{Pressure } 0.14 \cos \frac{360^\circ}{24} (t - 4.0) \text{ mm.},$$

where t is the solar time in hours.

Owing to the additive property of harmonic coefficients, to correct for air mass absorption we subtract from the vector representing the observed 24-hour wave in cosmic-ray intensity that corresponding to the same harmonic in pressure after multiplying the latter by the absorption coefficient -0.23% per mm. But there still remains another correction to be made for pressure. We have to remove the effect on cosmic-ray intensity at ground-level of the change in height

of the meson-producing layer which is in phase with the barometric variation. If we assume that the relative pressure change, $\delta p/p$, at the height of 16 km is the same as that at the ground, the amplitude of the vertical motion of the layer would be only 1.5 m, and the cosmic-ray intensity at ground-level would change by less than 0.01%. If, as is believed, the 24-hour variation of the barometer, which has its minimum at 4 p.m., is mainly the result of lateral flow of air from warmer towards cooler regions of the atmosphere, there is no reason to expect greater relative pressure change at 16 km. than at the ground. Quite probably, therefore, the 24-hour wave in cosmic-ray intensity after correcting completely

for pressure would be $0.22 \cos \frac{360^\circ}{24} (t - 17.3)^\circ$

With regard to the effect of daily temperature change, the examination of upper-air data obtained in England in the last few years by sending up sounding balloons every six hours shows that the height of the 75-mm. pressure-level is, on the average, appreciably greater at noon than at other hours. This is to be expected, since it is known that the temperature of the lower stratosphere is mainly controlled by radiation from the earth. We must clearly have a 24-hour wave due to temperature with the minimum at about noon in the observed variation of cosmic-ray intensity. I am not in a position to-day to give you the exact average value of the amplitude of this wave, but it seems to be of the order of 0.1% or rather less. It is easy to see that after correcting for this temperature effect, the maximum of the final 24-hour wave in cosmic-ray intensity will occur earlier in the afternoon and be greater than the observed maximum.

In order to determine the influence of the position of the earth in its orbit, the data have been averaged for the four-monthly seasons, December solstice, June solstice, Equinoxes. Table 2 contains the results obtained for the amplitudes and times of maximum of the 24-hour wave corrected for pressure.

Table 2 24-hour wave in cosmic-ray intensity corrected for pressure

	Amplitude	Time of maximum
	%	h
Dec. solstice	0.22	15.9
Equinoxes	0.18	16.7
June solstice	0.32	18.5

As the table shows, the time of the maximum is gradually delayed as we pass from cooler to warmer months and the effect of temperature becomes greater. Had this effect been eliminated, it is quite probable that the time of the maximum would have been the same and would have been nearer to noon for the three seasons, in keeping with the theoretical result of Vallarta and Godart. Another interesting feature is that the amplitude at the solstices is greater than at the equinoxes, in qualitative agreement with the calculations of Jánossy and Lockett, already quoted.

As, moreover, the gamma radiation is not effective in producing coincidences in my apparatus, these results for the 24-hour wave, after correction for meteorological influences, may be taken as supporting the view that one part of the diurnal

variation of cosmic-ray intensity is due to the effect on the primary particles of a magnetic field at the surface of the sun. The possibility, however, of explaining the 24-hour wave in cosmic-ray intensity as an effect of the earth's magnetic field, as suggested by Gunn (1932), cannot be excluded.

The harmonic analysis of the triennial curve reveals the existence also of a physically significant 12-hour wave. This wave is given below, together with the semi-diurnal barometric oscillation from my barograph readings.

$$\text{Cosmic rays} \quad 0.18 \cos \frac{360^\circ}{12} (t - 15.2)^\circ ;$$

$$\text{Pressure} \quad 0.22 \cos \frac{360^\circ}{12} (t - 10.2) \text{ mm.}$$

As we see, they are within one hour of being opposite in phase.

As before, to remove from the cosmic-ray wave the total effect of pressure we should have to know the amplitude of the vertical motion of the meson-producing layer, which is in phase with the barometric oscillation. If it is assumed that the relative pressure change at the 16 km level is the same as at the ground, that amplitude would not be greater than 2.5 m., but if we take into account that the solar semi-diurnal oscillation of the atmosphere is partly tidal in character, then the vertical displacement with which we are concerned might prove to be greater.

I may perhaps remind you that this oscillation, which appears all over the world with a constant phase and a regular variation of amplitude with latitude, seems to be excited tidally and thermally by the sun. Its large magnitude is considered to be due, as Kelvin suggested, to magnification by resonance, the atmosphere having a natural mode of free oscillation of period very near 12 solar hours, as shown by Pekeris (1937), when a certain vertical distribution of air temperature is assumed.

In developing this theory, Pekeris finds that the vertical motion of air particles due to the solar barometric oscillation is in phase at all levels with the change of pressure at the ground. Later on, he calculated the magnitude of the vertical displacement of an air particle in the ionosphere at the equator (see Appleton and Weekes, 1939) and found that at 125 km. it should be of the order of 1.45 km. The equivalent displacement for an air particle in the ionosphere over London should be, taking into account the variation in barometric oscillation with latitude, 323 m.

Experimental evidence on the possible magnitude of the increase of the vertical motion of air particles with height is provided by the results of Appleton and Weekes (1939) relating to variation with time of the equivalent height of the Kennelly-Heaviside layer. They found a lunar tide in this layer which is in phase with the lunar barometric oscillation at the ground, as found first by Chapman, and 7000 times greater, the amplitude of the tide in the Kennelly-Heaviside layer being of the order of 1 km.

It seems reasonable therefore to attribute to the layer at 16 km. a vertical displacement of amplitude greater than 2.5 m. In order to form an idea of the effect of a greater motion of this layer on the cosmic-ray intensity at the ground, let us assign to it an amplitude of only 22 m. in contrast to the 323 m. found by

Pekeris for the ionosphere. Then, by applying our value for the rate of decay of mesons, and allowing for the air mass absorption, we should find that the semi-diurnal variation of cosmic-ray intensity would be entirely accounted for by the motion of the meson-producing layer due to the barometric oscillation. Assuming this is so, if the 12-hour wave in cosmic rays were merely an effect of the 12-hour wave in atmospheric pressure, we should expect the seasonal changes of the two semi-diurnal variations to follow each other closely. Now this is in fact what my results reveal. Table 3 enables us to compare the amplitudes and times of second maximum of both waves for the solstices and the equinoxes.

Table 3

	Amplitude		Time of second max	
	<i>P</i> mm	<i>C R</i> %	<i>P</i> h	<i>C R</i> h
Solstices	0 21	0 16	10 2	15 4
Equinoxes	0 25	0.22	10 1	14 7

As the table shows, the amplitudes, within the limits of accuracy of the observations, are closely correlated and the phases remain constant.

It may be that mesons are generated at a greater height than 16 km, as suggested by the results of the experiments of Schein, Jesse and Wollan (1941). If this is so, it would be an additional reason for concluding that the semi-diurnal variation of cosmic-ray intensity at ground-level can represent the barometric oscillation in the region of the atmosphere where mesons are formed. In this connexion it would be interesting to see whether the moon has an appreciable influence on cosmic rays.

As the barometric oscillation is greater at lower latitudes, we must expect the magnitude of the semi-diurnal change of cosmic radiation to vary in the same direction, though the longer lifetime of mesons as we approach the equator tends to offset the effect.

Unfortunately I cannot compare my results with those of other workers, since it has been the practice to analyse the data after correcting them for pressure by using the so-called barometric coefficient.

§4 WORLD-WIDE CHANGES

In the last decade many investigators have reported certain effects on cosmic-ray intensity during great magnetic storms. The most notable of these effects is a decrease of a few per cent in the cosmic-ray intensity which is concomitant with a decrease in the horizontal component of the earth's magnetic field.

The magnetic storm of 1 March 1942 provides a most remarkable example on account of the exceptional magnitude of the decrease and other features of the storm effect on cosmic rays. At that time cosmic-ray recorders were in operation in various widely separated parts of the world, and a comparison of the results obtained reveals the world-wide character of the phenomenon. Figure 3 illustrates the changes of the bi-hourly mean number of cosmic rays expressed in percentages of the pre-storm value, as observed in London. The data, reduced to constant atmospheric pressure, were obtained with the coincidence arrangement mentioned previously. The diagram indicates that the regular decrease began

shortly before the onset of the magnetic storm, and reached the unusually high value of about 11%. After the first rapid recovery of about half the initial drop, which nearly coincided with the end of the storm, after midnight, the recovery was very slow, probably owing to the various geomagnetic disturbances which

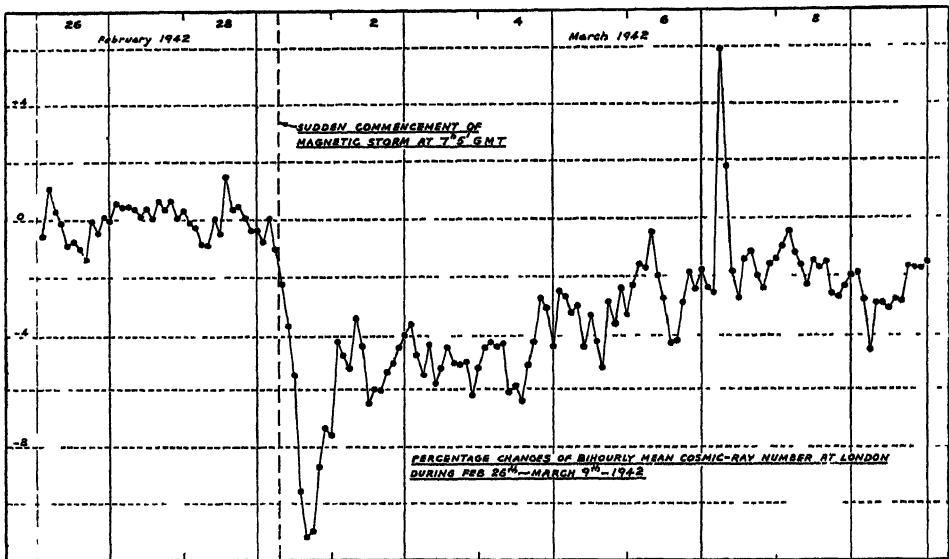


Figure 3.

This diagram differs from that published at the time (1942) in that bi-hourly means are used now to facilitate comparison with the diagrams published later

took place during the following days. Other features worth noticing are the increase on 28 February and the much greater one on 7 March. Increases as large as that of 7 March seem to be quite exceptional.

Figure 4 shows the diagram, as given in a note by Lange and Forbush (1942), of bi-hourly changes in cosmic-ray intensity at Godhavn (Greenland) in geomagnetic latitude 80° N., Cheltenham (Maryland) in geomagnetic latitude 50° N., and Huancayo (Peru) in geomagnetic latitude 0.6° S. during the same magnetic storm. At each of these three stations the data were obtained with Compton-Bennett meters protected by a 10-cm. lead shield. [In the same note, another diagram is given representing the bi-hourly changes, during the same storm, at Christchurch (New Zealand), and this reveals precisely the same features.] The very striking similarity of the simultaneous changes in cosmic-ray intensity at the three stations and London proves the world-wide character of the phenomenon beyond doubt. Apart from the smaller value of the increase on 28 February in London, the only outstanding discrepancy is the non-occurrence at Huancayo of any of the increases on 28 February and 7 March, which took place simultaneously at all the other stations. This particular result seems to suggest that the new particles responsible for both increases were unable to reach the magnetic equator, and, therefore, their energy was within the energy range which is sensitive to the magnetic field of the earth. Instead, the magnitude of the drop, within the limits of accuracy of bi-hourly means, is roughly the same at all stations, though owing to the fact that the altitude of Huancayo is 3350 metres

and that of the other stations is less than 100 it is not possible to say whether any variation of the drop with latitude exists.

The solar phenomena which preceded the onset of this storm were in some respects exceptional. As reported by Newton (1942), a great sunspot crossed

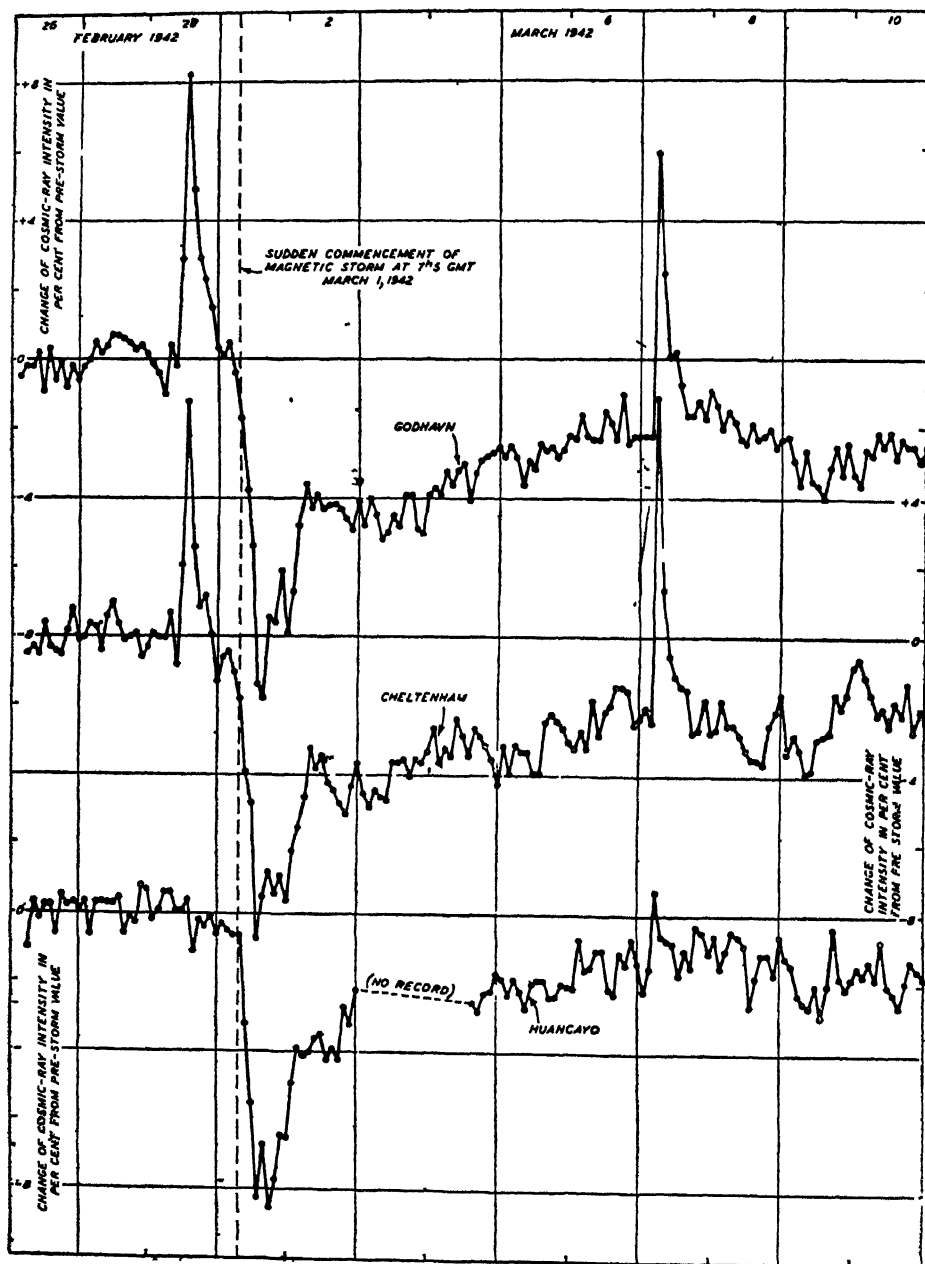


Figure 4.

the solar disk between 22 February and 7 March. The maximum area of the spot was 2000 millionths of the sun's hemisphere. A very extensive and brilliant eruption was observed over the umbra of the great sunspot nineteen and a half

hours before the beginning of the storm. This eruption appears to have been of exceptionally long duration.

Judging, however, from the magnitude of the range in the horizontal magnetic force, H , of 512 γ , the magnetic storm was not as intense as some others which have been accompanied by a much less violent disturbance in cosmic-ray intensity. To give an example, the range in H of the great magnetic storm of 4 July 1941 was, from the Abinger traces, about twice the range of the storm of 1 March 1942. Nevertheless, no appreciable effect on cosmic-ray intensity was shown by the records obtained in London for the 40-hour period following the abrupt commencement of the storm at 3^h 42^m a.m. of 4 July. Only at the end of this period, when the storm was over, did a decrease begin, and even then it was very small (less than 3%). But as another, though moderate, storm occurred a few hours later, on 6 July, it is possible that this small decrease was associated with the second disturbance rather than with the first one, and if so, the conclusion would be that the great storm of 4 July had no effect on cosmic radiation. Forbush (1938) has reported the case of the magnetic storm of 21 August 1937, during which no effect was detectable in the cosmic-ray records of Cheltenham (U.S.A.), Teoloyucan (Mexico) or Huancayo (Peru). However, all observers agree that the ratio of the relative change in cosmic-ray intensity to that of the horizontal component, while being always positive, differs in different storms.

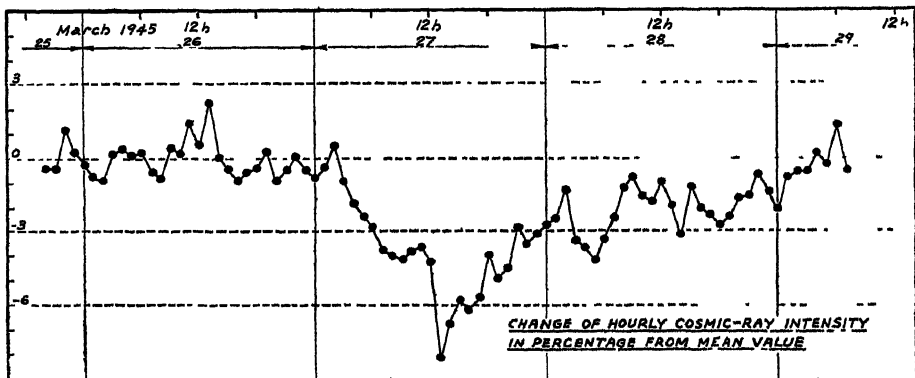


Figure 5.

When smaller changes in cosmic-ray intensity which cannot be explained as being due to meteorological influences are considered, the observations made in London show that they are also more or less closely associated with variations of the geomagnetic field. These fluctuations generally occur simultaneously with, or within a short time of, the fluctuations in the magnetic field of the earth, but as in the case of magnetic storms, the two variations may be of quite different magnitudes. The magnitude of the changes in the daily means of cosmic-ray intensity associated with these minor geomagnetic disturbances only rarely proves to be greater than 2% of the mean value. In this regard, however, a decrease of a very exceptional order of magnitude was observed in London on 27 March 1945. Figure 5 shows, in percentages of the mean value, the hourly change of cosmic-ray intensity, reduced to constant pressure, for a period of a few days around 27 March 1945. Within the limits of accuracy of hourly observations, the decrease of about 8% indicated by the diagram, which was not

associated with a magnetic storm, is, as far as I know, unprecedented in all records of cosmic rays. According to information given by the Royal Observatory, nothing abnormal was observed either in the magnetic activity or in the solar phenomena during that period, although a comparatively slight magnetic disturbance was recorded at Abinger, chiefly on the day preceding that of the decrease in cosmic-ray intensity. A sunspot, but only of the moderate area of 750 millionth of the sun's hemisphere, crossed the central meridian on the 27th after midnight. The complexity of these phenomena is evident.

These fluctuations in cosmic-ray intensity associated with magnetic or solar activity should be considered as taking place simultaneously all over the world. Striking evidence that the world-wide character of variations in cosmic radiation is not confined to periods of great magnetic storms has been given by Forbush (1939). He made the analysis of the data obtained at five widely separated stations and found that after eliminating a seasonal wave of a 12-monthly period which appeared at each of the stations except at Huancayo, there existed a very close correspondence between the remaining curves for all stations.

In addition, we have the finding, first reported by Hess and Graziadei (1936), of a 27-day periodicity in cosmic-ray intensity associated with the mean period of the sun's rotation relative to the earth. These results, therefore, can be taken as giving strong evidence that all appreciable changes at ground-level in cosmic radiation other than meteorological are of a world-wide character.

To account for these world-wide variations, some investigators, following Chapman (1937), think they may result from the formation or fundamental alteration of a system of westward currents concentric with the earth flowing in the high ionosphere or in the outer space, due to the emission of electric particles from the sun at the time of a geomagnetic disturbance. The magnetic field of this current-system superimposed on the earth's dipole would give rise to a decrease of the magnetic force in the regions inside and to an increase outside the current. Then the number of incoming particles reaching the earth might be reduced, according to the theory of Lemaitre and Vallarta, and the main phase of the phenomenon, particularly in the case of magnetic storms, would be explained.

If by means of this theory it is possible to account for the diversity of phenomena observed, we shall have to wait, as Forbush has pointed out, until the difficult problem of determining the paths of cosmic-ray particles in the field of the earth's dipole and the system of westward currents has been solved.

There is no doubt that the relation between cosmic rays and magnetic activity presents a very complicated problem. But I am convinced we shall not be discouraged by the difficulties involved in its solution. For when we have solved them we shall know more not only about cosmic rays, but also about the complex phenomena of geomagnetism.

REFERENCES

- APPLETON and WEEKS, 1939. *Proc. Roy. Soc. A*, 171, 171.
 BEARDSLEY, 1940. *Phys. Rev.* 57, 336.
 BENEDETTO, ALTMANN and HESS, 1942. *Phys. Rev.* 61, 266.
 BLACKETT, 1938. *Phys. Rev.* 54, 973.
 CHAPMAN, 1937. *Nature, Lond.*, 140, 423.
 DUPERIER, 1941. *Proc. Roy. Soc. A*, 177, 204.
 DUPERIER, 1942. *Nature, Lond.*, 149, 579.

- DUPERIER, 1944 *Terr Mag* 49, 1
 DUPERIER, 1944 *Nature, Lond*, 153, 529.
 EHMERT, 1937 *Z Phys* 106, 751
 FORBUSH, 1937 *Terr Mag.* 42, 16; 1938. 43, 203.
 FORBUSH, 1939. *Rev Mod Phys* 11, 168.
 GUNN, 1932 *Phys Rev* 41, 683
 HESS, 1936. *Terr Mag* 41, 9 and 345
 HESS and BENEDETTO, 1941 *Phys Rev* 60, 610
 JÁNOSSY and LOCKETT, 1941 *Proc Roy Soc A*, 178, 52
 LANGE and FORBUSH, 1942 *Terr Mag* 47, 331
 NEWTON, 1942 *The Observatory*, 64, 260
 PEKERIS, 1937 *Proc Roy Soc A*, 158, 650
 SCHEIN, JESSE and WOLLAN, 1941 *Phys Rev* 59, 615
 THOMPSON, 1938 *Phys Rev* 54, 93.
 VALLARTA and GODART, 1939 *Rev Mod. Phys* 11, 180

THE THERMAL EXPANSION OF GRAPHITE FROM 15° C. TO 800° C.: PART I. EXPERIMENTAL

BY J. B. NELSON AND D. P. RILEY,

Cavendish Laboratory, Cambridge

(Now at British Coal Utilisation Research Association)

MS received 23 March 1945

ABSTRACT The variation with temperature of the a and c unit-cell dimensions of hexagonal Ceylon graphite has been measured over the temperature range 15°–800° C. by the x-ray powder method. At 14° 6 C.,

$$a = 2.4562 \pm 0.0001 \text{ kx} \quad c = 6.6943 \pm 0.0007 \text{ kx}$$

The carbon-carbon bond length,

$$\text{C}-\text{C} = 1.4210 \pm 0.0001 \text{ \AA}.$$

The a dimension shows a slight contraction up to about 400° C., a small expansion occurring above this temperature. The thermal expansion in the c direction is large; the average value for α over the temperature range is 28.3×10^{-6} . The complex nature of the expansion in both directions is discussed qualitatively.

§ 1. INTRODUCTION

A STUDY of the thermal expansion of graphite is of interest for two main reasons. The first is the need for extending our knowledge of the solid state in the field of anisotropic crystals; the second is the importance of precise information concerning the effect of heat on elementary carbon in connection with the technological processes of carbonization and coking. The x-ray method is the only suitable one for an investigation of this character. Other methods, the results of which are summarized in § 2, are not able to do more than measure the *mean* linear expansion of the crystal. Such information is of little significance thermodynamically, the essence of the problem being the *anisotropy* of the structure. The work reported here shows how valuable a tool x-ray

diffraction methods can be in measuring the effect of temperature on such structures

§2 PREVIOUS WORK

Previous x-ray work comparable to ours has been carried out by Backhurst (1923) in the course of an investigation of the variation in intensity of x-ray reflections with temperature. He measured with the ionisation spectrometer and molybdenum radiation the shift of the 0004 reflection (called by him 0001) on heating from 17° to 870° c. In this way an average value over this range of 26.7×10^{-6} was derived for $\alpha_{||}$, the coefficient of linear expansion in the *c* direction.* This value is somewhat lower than the average value of 28.3×10^{-6} found by us. He did not measure the coefficient in the *a* direction, α_{\perp} , but surmised that a contraction and not an expansion might occur.

Hirata (1931) has investigated the behaviour of carbon in an incandescent arc at a temperature of about 4000° c. Even at this temperature, no measurable expansion in the *a* direction occurred, but by observing the shift of the 0002 reflection (with copper radiation) he derived these average values for $\alpha_{||}$ over this large temperature range:

$\alpha_{||} = 39.1 \times 10^{-6}$ assuming a temperature of 4200° c. by measurements of surface brightness.

$\alpha_{||} = 45.1 \times 10^{-6}$ assuming an average temperature of 3700° c.

The thermal expansion of graphite has also been measured by various other methods, and the following values have been recorded.

	Expansion coefficient	Temperature (° c)
Fizeau (1869)	7.86×10^{-6}	40°
	7.96×10^{-6}	50°
Muraoka (1881)	3.8×10^{-6}	26° to 302°
Day and Sosman (1912)	0.55×10^{-6} to 3.2×10^{-6}	0° to 1800°
Hidnert and Sweeney (1927)	2.7×10^{-6}	20° to 600°
	3.7×10^{-6}	
Erfling (1939)	2.34×10^{-6}	-195° to -183°
	6.695×10^{-6}	+20° to +40°
Dewar (1902)	7.33×10^{-6} (vol. coeff.)	-190° to +17°

In no case was a single crystal of graphite used. As the specimens probably consisted of aggregations of small graphite crystals variously oriented, the values can only refer to the mean coefficient of linear expansion $\bar{\alpha}$. In a hexagonal crystal

$$\bar{\alpha} = \frac{2}{3}\alpha_{\perp} + \frac{1}{3}\alpha_{||}$$

and, using our data, $\bar{\alpha} \approx 8 \times 10^{-6}$ at 20° c. Several of the values are much too low and probably refer to the expansion of the aggregated solid as a whole rather than to the expansion of the individual crystals. It is interesting that Fizeau,

* The symbols $\alpha_{||}$ and α_{\perp} refer to the coefficients of linear expansion parallel and perpendicular to the hexagonal axis.

who was the first to measure the thermal expansion of graphite in the course of his classical researches, was also responsible for the best value of $\bar{\alpha}$.

The work of Erfling requires special comment. Employing an optical method, he attempted to measure α_{\perp} by taking care to orient the crystals in the specimen. In spite of his precautions Erfling must have been measuring the mean coefficient of expansion. His value of 6.695×10^{-6} at 20° to 40° C. is probably a low value for $\bar{\alpha}$; it is much too large to be α_{\perp} , which we have found to be a small negative quantity over this range.

§ 3. APPARATUS

The camera used was the high-temperature 19 cm.-diameter powder-camera described by Wilson (1941). The camera angle θ_k for this instrument has been carefully determined by Wilson and Lipson (1941). The use of this camera in the determination of accurate unit-cell dimensions had been previously investigated (Nelson and Riley, 1945). The method of measuring the temperature of the specimen was essentially that of Wilson. We found, however, that the melting point of aluminium filings did not provide a satisfactory high-temperature calibration as the aluminium reacted with the silica capillary in which it was contained. Antimony (Hilger H.S.), with a melting point of 630.5° C., was used in its place. It can be said that the error in temperature measurement did not exceed $\pm 1^{\circ}$ C. at the highest temperature investigated and was certainly less than this at the lower temperatures.

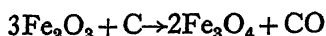
§ 4. PREPARATION OF THE SPECIMEN

Some difficulty was encountered in obtaining a specimen of graphite which combined high purity with the ability to give the sharp diffraction lines necessary for accurate measurements.

An examination of powder photographs of available graphites ("Kish", Acheson, Mexican, Bavarian, Travancore, Ceylon) showed that the Ceylon specimen gave the sharpest lines, although the purity was not of the highest (4% ash). It was felt, however, that purification by the usual chemical treatments (boiling with NaOH, H_2SO_4 or HF solution) was to be avoided because of the ease with which graphite forms complexes which show an appreciably altered c dimension. The Ceylon graphite was therefore chosen in spite of its low purity.

After reducing the material to pass a 300-mesh B.S.S. sieve, several silica capillary tubes of bore approximately equal to 0.05 cm. and a wall thickness of 0.003 cm. were filled. It was thought advisable to remove traces of adsorbed gases by heating to about 850° C. during the evacuation of the filled capillaries. However, as soon as even gentle heating was applied to them, the graphite powder was completely swept out, presumably by gases released during the heating. This made it necessary to heat the powdered graphite in bulk in a large-bore (1.5 cm.) continuously evacuated silica tube. Heating for 30 minutes at approximately 850° C. was found to be an ample treatment for completely removing the gas. Photographs of the graphite taken in an ordinary powder camera showed that the only extra lines discernible, besides those due to a small amount of the rhombohedral modification of graphite, were those of Fe_3O_4 in the heat-

treated material and hydrated Fe_2O_3 in the untreated. It seems fairly certain, therefore, that the evolution of gas is connected with the reaction



together with the evolution of water vapour.

No trouble was encountered in evacuating, heating and sealing off the capillaries after heat-treatment. The sealed-off capillary with the smallest bore and thinnest, most uniform wall thickness was selected for the exposures, which were made in order of descending temperature. Its bore, as measured by a microscope ocular micrometer, was 0.048 cm. and the wall thickness 0.0025 cm.

Subsequent precision measurements of both the a and c dimensions of the heat-treated and untreated graphites at room temperature were in excellent agreement. It would appear, therefore, that the impurities present (chiefly oxide of iron) have no influence on the measurements as they simply form a mixture with the graphite itself.

§ 5. THE X-RAY PHOTOGRAPHS

X-ray photographs were taken at the following temperatures: 14°C , 150° , 300° , 450° , 600° , 700° , 800°C .

The temperature fluctuation during the exposure was greater the higher the temperature, but in no case did it exceed $\pm 0^\circ 5\text{C}$. The photographs were taken with unfiltered manganese radiation with exposures of about 200 ma. hrs.

The high-angle portions of the x-ray photographs are reproduced in figure 1. Even a qualitative comparison leads to some interesting conclusions. The $hki0$ reflections ($11\bar{2}0$ and $20\bar{2}0$) do not move with change of temperature, while the 0006 doublet moves towards the low-angle end of the film quite markedly with increasing temperature. It can also be observed with the $hki1$ reflections that as the l index increases ($11\bar{2}2$, $11\bar{2}4$, $10\bar{1}5$), the rate of line shift with temperature also increases. The thermal expansion of the lattice is clearly anisotropic, the expansion occurring along the c axis. It should be noted that, at 15°C ., the 553 doublet of Fe_3O_4 is practically coincident with the $11\bar{2}4$ β line of graphite. Resolution is effected at higher temperatures due to the different thermal expansions.

Another effect which is clearly visible concerns the relative intensities of the diffraction lines. The relative intensity of the 0006 reflections compared with the intensities of the $hki0$ reflections ($11\bar{2}0$, $20\bar{2}0$) shows a progressive diminution with temperature. This fading of the 0006 line is due to the large amplitude of thermal vibrations perpendicular to the layers, and obviously is in agreement with the pronounced anisotropy of thermal expansion. The $10\bar{1}5$ reflection shows a similar effect because of the high l index.

There are also nine extra lines due to the Fe_3O_4 previously mentioned, although only one (553 doublet) can be seen in figure 1. The identity of this compound was established by comparison of the interplanar spacings and intensities of the lines observed with that of pure Fe_3O_4 . To obtain the true cell dimensions, the calculated apparent length of the cubic unit cell was plotted against the extrapolation function $\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta}$ for the four strongest lines on both the room

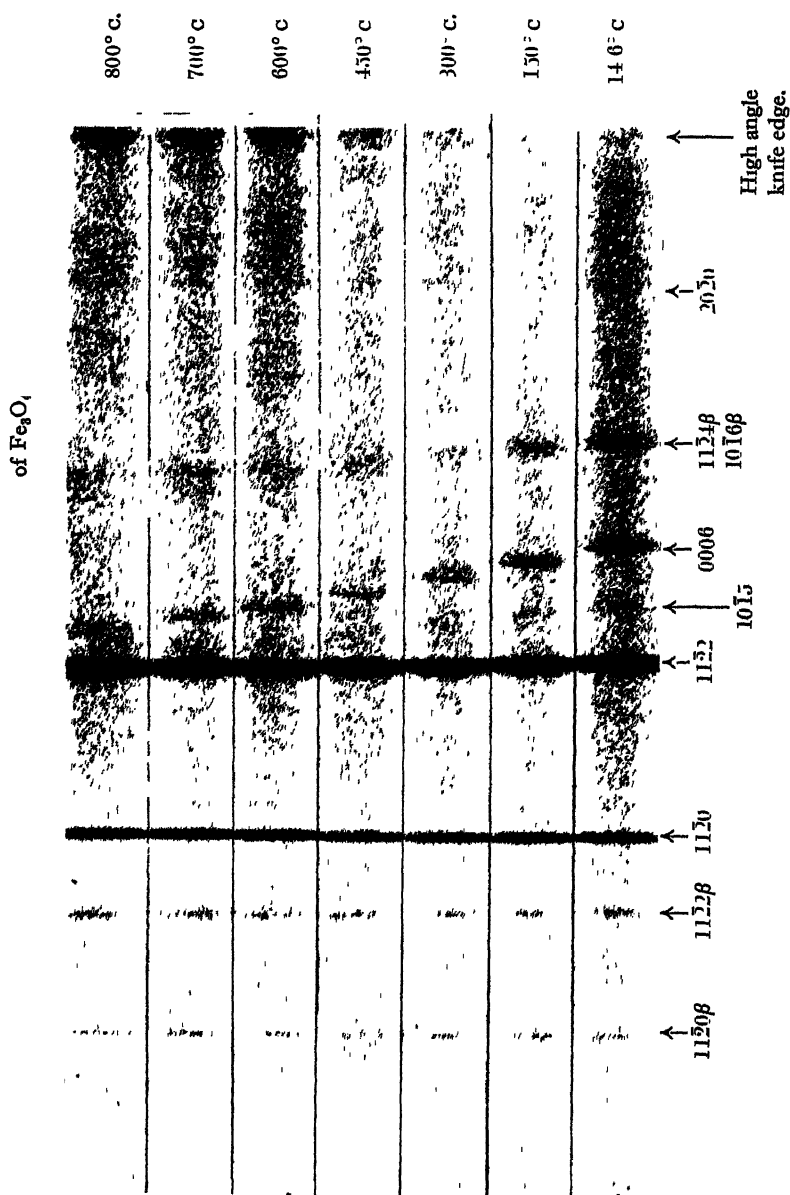


Figure 1 X-ray photographs of graphite at various temperatures.

temperature and 800° c. films The results are as follows :

Temperature (°C.)	<i>a</i> (kx.)
15	8 378
800	8 475
20 (pure Fe ₃ O ₄ , Wyckoff and Crittenden 1925)	8 37±0.01

Mean linear coefficient of thermal expansion, 14.7×10^{-6} (range 15° to 800° c)

§ 6. MEASUREMENT OF UNIT-CELL DIMENSIONS AND DERIVATION OF EXPANSION COEFFICIENTS

The best radiation to use (manganese) gives the 2020 α -doublet at $\theta = 80^\circ$ to 81° and the 0006 α -doublet at $\theta \approx 70^\circ$ at room temperature. These angles, and particularly the latter one, are not as high as is desirable in precision measurements, but it is impossible to effect any improvement in this case by change of x-ray wave-length. Hence, particular attention had to be paid to the extrapolation procedure adopted.

Work described elsewhere (Nelson and Riley, 1945) has shown that a reliable linear extrapolation is obtained down to $\theta \approx 30^\circ$ if the unit cell dimension under consideration is plotted against $\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta}$. The apparent *c* dimensions given by the 0004 ($\theta = 38^\circ.9$ at room temperature) and 0006 reflections were therefore plotted against this function. A linear extrapolation to $f(\theta) = 0$ gave the *c* value quoted for each temperature. As the extrapolation lies through only two points, the accuracy claimed is mainly dependent on the suitability of the function of θ used in the plot. We have assigned an error of about one part in 10,000 to the *c* values derived in this way. The 0004 lines were accompanied by a weak satellite line on all the films. This feature was shown to arise from a preferred orientation of the graphite crystals in the specimen. The effect of this on the accuracy of the determination of the *c* dimension has already been discussed (Nelson and Riley, 1945).

In order to obtain the best value for the *a* dimensions at each temperature, the apparent *a* values derived from the following lines were plotted against $\frac{\cos^2 \theta}{\sin \theta} + \frac{\cos^2 \theta}{\theta}$:

$$\begin{aligned} &2020 \alpha_1, \alpha_2. \\ &11\bar{2}0 \alpha_1, \alpha_2, \beta. \\ &10\bar{1}0 \alpha, \beta. \end{aligned}$$

At room temperature, the θ values for the 1010 α and β lines were roughly $29^\circ.6$ and $26^\circ.6$ respectively. We did not assign them much weight in the extrapolation, but they serve to confirm its general trend. The error assigned to any given *a* dimension is approximately one part in 25,000. The refraction correction for *a* or *c* is negligible in comparison with the probable experimental error.

The values derived for the *c* dimensions at the various temperatures are given in table 1. A plot of *c* against temperature gives a smooth curve deviating only

slightly from linearity. An expression in t and t^2 was adopted to represent this curve, and the results were analyzed by the method of least squares. The best-fitting relation was found to be

$$c = 6.6915_4 + 180.70 \times 10^{-6}t + 12.63 \times 10^{-9}t^2, \quad \dots\dots(1)$$

where t is expressed in $^{\circ}\text{C}$. and c in kx .

Table 1. c dimensions of graphite at various temperatures, in kx

Temperature ($^{\circ}\text{C}$)	Observed	Calculated	Difference (calc.-obs.)
14.6	6.6943	6.6941 ₈	-1.2×10^{-4}
150	6.7185	6.7189 ₃	+4.3
300	6.7474	6.7468 ₈	-5.2
450	6.7752	6.7754 ₁	+2.1
600	6.8046	6.8045 ₁	-0.9
700	6.8240	6.8242 ₂	+2.2
800	6.8443	6.8441 ₈	-1.2

The values of c calculated from this equation are also given in the table. The deviations between the calculated and observed values are too large for equation (1) to be considered a satisfactory expression. The mean deviation is 2.4×10^{-4} , and c is therefore expressed with an average accuracy of only one part in 2800, whereas the observed values themselves are accurate to about one part in 10,000. That the assumption of a parabolic expression for c is not justified is confirmed by examining the variation with temperature of the expansion coefficient. Equation (1) leads to the following linear expression for the coefficient

$$\alpha_{||} = \frac{1}{c} \cdot \frac{\partial c}{\partial t}; \quad \alpha_{||} = 27.00 \times 10^{-6} + 3.05 \times 10^{-9}t \quad \dots\dots(2)$$

The straight line of equation (2) is compared with values of $\alpha_{||}$ derived directly from the experimental data in figure 2.

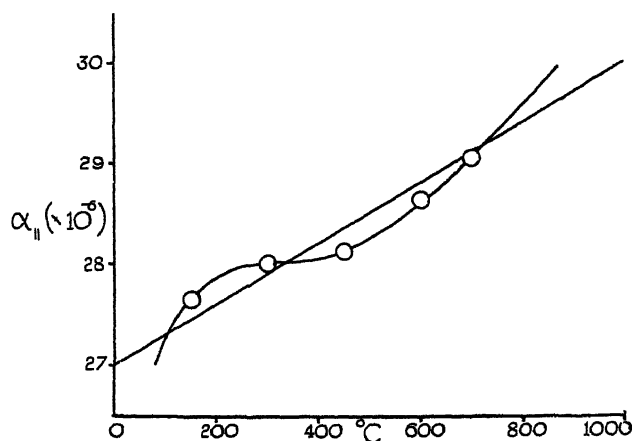


Figure 2 Comparison of observed $\alpha_{||}$ values with the best linear representation.

These values of $\alpha_{||}$ were derived in the following way. From the values c_1, c_2, c_3 of the c dimension at neighbouring temperatures t_1, t_2, t_3 , two extreme

values for $\frac{\Delta c}{\Delta t}$ at the middle temperature t_2 were obtained. The mean of these, divided by c_2 , was taken as the best value of $\alpha_{||} = \frac{1}{c} \cdot \frac{\Delta c}{\Delta t}$ at the temperature t_2 . Owing to the large values of Δt involved, this derivation of $\alpha_{||}$ cannot be very accurate and, in particular, it is impossible to judge its course at either end of the temperature range.

It would appear from figure 2, however, that the temperature variation of $\alpha_{||}$ is of a complex form and is only approximately represented by a linear expression in t . A more exact representation is derived on theoretical grounds in Part II. The average value for $\alpha_{||}$ over the temperature range is 28.3×10^{-6} .

The measurements of the a dimension at the various temperatures showed that a shrinkage followed by an expansion occurred. In order to check this, the films were measured independently by each of us, and the best extrapolated a values are given in table 2.

Table 2

Temperature (°C.)	a dimensions in kx.	
	Measurement A	Measurement B
14.6	2 45626	2.45617
150	2 45575	2.45564
300	2 45570	2.45563
450	2 45577	2.45571
600	2.45574	2.45574
700	2.45607	2.45602
800	2.45613	2.45619

It will be seen that the agreement between the two sets of measurements is good. Two further checks were carried out in order to confirm the general nature of the effect.

(a) In measurement A the θ values for 2020 α_1 at 15° and 300° c. are 80°·419 and 80°·499 respectively. The difference is therefore 0°·080 which corresponds to a line shift of 0.27 mm. This distance lies outside the limits of error in observing line positions (± 0.10 mm.). Therefore we conclude that there is a real change in angle and a real diminution in the a dimension.

(b) The measured θ values for the 2020 and 1120 lines were plotted against temperature for both sets of measurements, A and B. A general rise and then fall in θ was observed, in accordance with the shrinkage and subsequent expansion of a . This showed that the extrapolations could not be systematically in error.

Rough values of the thermal contraction and thermal expansion coefficients can be derived from table 2, and are :

Temperature range	α_{\perp}
0° to 150° c.	-1.5×10^{-6}
ca. 400° c.	0
600° to 800° c.	$+0.9 \times 10^{-6}$

The a dimension first contracts and then expands. It reaches its lowest value of approximately 2 4557 kx. at a temperature of about 400° c.

The unit cell dimensions at 15° c. are

$$a = 2.4562 \pm 0.0001 \text{ kx} ; \quad c = 6.6943 \pm 0.0007 \text{ kx.}$$

and the carbon-carbon bond length, expressed in Å, is

$$\text{C—C} = 1.4210 \pm 0.0001.$$

We found no evidence that the rhombohedral modification of graphite (Lipson and Stokes, 1942) has a different thermal expansion from that of the hexagonal variety. Our measurements, however, refer specifically to the latter.

§7. DISCUSSION

A qualitative explanation of the above results will be given first, in which the treatment of Gruneisen and Goens (1924) (see also Roberts (1940)) of the thermal expansion of hexagonal crystals will be followed.

Graphite is a perfect example of a layer structure. Each layer consists of a very large number of carbon atoms covalently linked to form what may be considered to be a huge polynuclear aromatic macromolecule. Each layer is planar, or very nearly so. The carbon atoms within a layer are very strongly linked, the C—C distance being about 1.42 Å. The bonding between adjacent layers, however, is weak, as is indicated by the large interlayer spacing of about 3.35 Å. It is probable that the interlayer forces are mainly of a van der Waals type, although they may not be entirely so. The graphite structure is therefore one of great anisotropy, and this will be reflected in all its directional physical properties.

The elastic deformability parallel to the principal Z axis will be much larger than that perpendicular to it owing to the tighter interatomic bonding within the layers. This may be written in terms of the elastic moduli as $s_{33} \gg s_{11}$.

Under the influence of temperature and of zero-point energy the atoms will vibrate, and the limiting frequency ν_{m_z} for vibrations parallel to the axis will be much smaller than for those perpendicular to it, ν_{m_x} . The solid will therefore possess two characteristic temperatures defined by these two limiting frequencies.

$$\Theta_z = \beta \nu_{m_z},$$

$$\Theta_x = \beta \nu_{m_x},$$

where

$$\beta = \frac{h}{k}$$

h = Planck's constant, k = Boltzmann's constant

and

$$\Theta_x > \Theta_z.$$

At low temperatures, quantum theory requires nearly all the thermal energy to be absorbed in oscillations parallel to the axis because of the low value of Θ_z as compared with Θ_x . These oscillations will, therefore, have much the greater amplitude and the thermal expansion parallel to the axis, $\alpha_{||}$, will be appreciable, whereas that perpendicular to the axis, α_{\perp} , will be negligible.

Thus the lattice is effectively being stretched in one direction only and a lateral contraction (Poisson contraction), proportional to the modulus s_{13} , will

accompany the stretching. At low temperatures, the lateral contraction will be greater than the small thermal expansion within the layers, and α_{\perp} will therefore be negative. As the temperature increases, more energy will go into vibrations perpendicular to the axis, until a temperature T_m is reached when the thermal expansion exactly balances the lateral contraction. At this temperature $\alpha_{\perp} = 0$. Above this temperature the thermal expansion within the layers will increasingly outweigh the lateral contraction, and α_{\perp} will be positive and increase with temperature. When this is the case an elastic contraction parallel to the axis ($\propto s_{13}$) will be contributed, which will have the effect of diminishing the rate of increase of α_{\parallel} and at a high enough temperature may cause α_{\parallel} to diminish after passing through a maximum. The apparent flattening of the curve for graphite between 250° and 450° c. (figure 2) can be explained in this way.

Grüneisen and Goens (1924) observed with zinc and cadmium similar effects to those for graphite but at lower temperatures. Zinc and cadmium are hexagonal layer structures but their anisotropy is not so pronounced as that of graphite. The data for the metal are compared with those for graphite in the following table.

		Θ_x	Θ_s
Cadmium	ca 45° K.	214	160
Zinc	„ 80° K.	320	200
Graphite	„ 660° K.	2280	760

The values of Θ for graphite are those given by Magnus (1923).

It will be noted that Θ_x for graphite is very much higher than for Zn or Cd as is the ratio Θ_x/Θ_s . It would therefore be expected that T_m also would be very much higher. Grüneisen and Goens themselves suggest, when briefly discussing the implications of their work in the case of graphite, that this substance would have α_{\perp} negative up to higher temperatures than does Zn or Cd.

§8 ACKNOWLEDGMENTS

We wish to thank Professor Sir Lawrence Bragg, F.R.S., for giving us hospitality and working facilities in the Cavendish Laboratory. The work constitutes part of a research programme of the British Coal Utilisation Research Association on the structure and properties of carbon and carbon derivatives, and we are grateful to Dr. D. H. Bangham for his encouragement and interest. Dr. A. R. Stokes gave us access to his collection of graphite samples and x-ray photographs, and we are indebted to him and to Dr. H. Lipson for several helpful discussions.

REFERENCES

- BACKHURST, I., 1923. *Proc. Roy. Soc. A*, **102**, 340.
 DAY, A. L. and SOSMAN, R. B., 1912 a. *J. Acad. Sci., Wash.*, June 19, 284; 1912 b. *Industr. Engng. Chem.*, **4**, 490.
 DEWAR, J., 1902. *Proc. Roy. Soc. A*, **70**, 237.
 ERFLING, H. D., 1939. *Ann. Phys., Lpz.*, **34**, 136.
 FIZEAU, H., 1869. *C.R. Acad. Sci., Paris*, **68**, 1125; *Ann. Phys., Lpz.*, **138**, 26.
 GRÜNEISEN, E. and GOENS, E., 1924. *Z. Phys.*, **29**, 141.

- HIDNERT, P and SWEENEY, W T, 1927 a. *Phys Rev* 29, 371; 1927 b *Tech Pap Bur Stand, Wash.*, 21, 223
 HIRATA, M., 1931. *Sci Papers Inst Phys. Chem Res, Tokyo*, 15, 219
 LIPSON, H. and STOKES, A R, 1942 *Proc Roy. Soc A*, 181, 101
 MAGNUS, A, 1923 *Ann Phys, Lpz*, 70, 303
 MURAOKA, H, 1881. *Ann Phys, Lpz*, 13, 307.
 NELSON, J. B and RILEY, D. P., 1945 *Proc Phys Soc* 57, 160
 ROBERTS, J K, 1940. *Heat and Thermodynamics*, 4th ed (Cambridge: University Press).
 WILSON, A. J C, 1941 *Proc. Phys Soc* 53, 235
 WILSON, A. J C and LIPSON, H., 1941 *Proc Phys Soc.* 53, 245.
 WYCKOFF, R W G. and CRITTENDEN, E. D., 1925. *J. Amer Chem. Soc.* 47, 2876.

THE THERMAL EXPANSION OF GRAPHITE: PART II. THEORETICAL

By D. P. RILEY,

Cavendish Laboratory, Cambridge

(Now at British Coal Utilisation Research Association)

MS received 23 March 1945

ABSTRACT. A theory of the thermal expansion of hexagonal crystals is derived, and shown to account quantitatively for the experimental data on graphite. Certain of the elastic moduli for graphite are estimated, and are at 18° c approximately:

$$\begin{aligned}s_{11} + s_{12} &= 1.8 \times 10^{-13} \text{ cm/dyne} \\ s_{13} &= -4.3 \times 10^{-13} \text{ , ,} \\ s_{33} &= 58.5 \times 10^{-13} \text{ , ,}\end{aligned}$$

Hence the two linear compressibilities are, at 18° c, approximately:

$$\kappa_{\perp} = -2.5 \times 10^{-13} \text{ cm}^2/\text{dyne}; \quad \kappa_{\parallel} = 50 \times 10^{-13} \text{ cm}^2/\text{dyne}.$$

§1 INTRODUCTION

GRÜNEISEN AND GOENS (1924) derived a theory of the thermal expansion of hexagonal crystals at low temperatures which enabled them to correlate the expansion data for zinc and cadmium with the elastic moduli and heat capacities. In order to achieve this for the experimental results for graphite described in Part I, a theory which is not restricted to low temperatures must be used. The object of this paper is to derive such a theory and apply it to the data for graphite. The approach is somewhat more straightforward than that of Grüneisen and Goens. In particular, the heat-capacity quantities used are more simply defined and are easily obtainable from the experimental data usually available.

§2. THERMAL EXPANSION OF HEXAGONAL CRYSTALS

In expressing thermal expansion, the basic thermodynamic relationships are:

$$(\partial T/\partial P)_P = -(\partial P/\partial T)_T$$

or

$$\left(\frac{\partial^2 V}{\partial T^2}\right)_P = -\frac{1}{T} \left(\frac{\partial C_p}{\partial P}\right)_T,$$

where V is the volume of the system, S the entropy, T the absolute temperature, and C_p the heat capacity at constant pressure.

Integration gives

$$V\alpha = \left(\frac{\partial V}{\partial T}\right)_P = - \int_0^T \left(\frac{\partial C_p}{\partial P}\right)_T \frac{dT}{T},$$

where α is the volume coefficient of expansion, the integration being performed at constant pressure.

Analogous expressions hold for the two linear expansion coefficients in a hexagonal crystal (Voigt, 1910). In this case, the x and y axes are equivalent and

$$\alpha_{\perp} = \alpha_x = \alpha_y = \frac{1}{l_x} \left(\frac{\partial l_x}{\partial T}\right)_P; \quad \alpha_{\parallel} = \alpha_z = \frac{1}{l_z} \left(\frac{\partial l_z}{\partial T}\right)_P,$$

where l_x and l_z are linear dimensions in the x and z directions, and α_{\perp} and α_{\parallel} are the linear expansion coefficients perpendicular and parallel to the hexagonal axis.

In the case of a monatomic solid, if v is the volume and S the entropy of a gram-atom, then

$$v\alpha_x = - \left(\frac{\partial S}{\partial X_x}\right)_T; \quad v\alpha_z = - \left(\frac{\partial S}{\partial Z_z}\right)_T,$$

where X_x is the pressure along x on a plane perpendicular to x , and where Z_z has a similar meaning. X_x and Z_z are sometimes called "thermal pressures" and are positive if the interior of the crystal is considered. Thus, since

$$S = \int_0^T \frac{C_p}{T} dT,$$

then
$$v\alpha_x = - \int_0^T \left(\frac{\partial C_p}{\partial X_x}\right)_T \frac{dT}{T}; \quad v\alpha_z = - \int_0^T \left(\frac{\partial C_p}{\partial Z_z}\right)_T \frac{dT}{T}, \quad \dots\dots (1)$$

where C_p is the heat capacity at constant pressure per gram-atom. $\frac{\partial C_p}{\partial X_x}$ is now written as

$$\frac{\partial \log l_x}{\partial X_x} \frac{\partial C_p}{\partial \log l_x} + \frac{\partial \log l_y}{\partial X_x} \frac{\partial C_p}{\partial \log l_y} + \frac{\partial \log l_z}{\partial X_x} \frac{\partial C_p}{\partial \log l_z},$$

and the elastic moduli* introduced by using the relations

$$\frac{\partial \log l_x}{\partial X_x} = -s_{11}; \quad \frac{\partial \log l_y}{\partial X_x} = -s_{12}; \quad \frac{\partial \log l_z}{\partial X_x} = -s_{13}.$$

Since for the hexagonal system

$$\frac{\partial C_p}{\partial \log l_y} = \frac{\partial C_p}{\partial \log l_x},$$

* The elastic moduli s_{11} , s_{12} , s_{13} are fourth-order tensors and can be written in full as s_{1111} , s_{1122} , s_{1133} , s_{2222} .

For a stress along X ,

s_{11} is the modulus giving the strain in the X direction,

s_{12} " " " " " Y "

s_{13} " " " " " Z "

For a stress along Z ,

s_{13} is the modulus giving the strain in the X direction,

s_{33} " " " " " Z "

In the discussion, s_{pq} are assumed to be independent of temperature

we have
$$-\frac{\partial C_p}{\partial X_x} = (s_{11} + s_{12}) \frac{\partial C_p}{\partial \log l_x} + s_{13} \frac{\partial C_p}{\partial \log l_z}$$

Similarly,
$$-\frac{\partial C_p}{\partial Z_z} = 2s_{13} \frac{\partial C_p}{\partial \log l_x} + s_{33} \frac{\partial C_p}{\partial \log l_z}.$$

These expressions can now be substituted in (1), and the two expansion coefficients written as

$$\alpha_x = (s_{11} + s_{12})q_x + s_{13}q_z, \quad \dots\dots(2a)$$

$$\alpha_z = 2s_{13}q_x + s_{33}q_z, \quad \dots\dots(2b)$$

where q_x and q_z are the so-called "thermal pressure coefficients" perpendicular and parallel to the hexagonal axis :

$$q_x = \frac{1}{v} \int_0^T \left(\frac{\partial C_p}{\partial \log l_x} \right)_T \frac{dT}{T}, \quad \dots\dots(3a)$$

$$q_z = \frac{1}{v} \int_0^T \left(\frac{\partial C_p}{\partial \log l_z} \right)_T \frac{dT}{T}. \quad \dots\dots(3b)$$

This represents the limit of purely thermodynamic reasoning. In order to evaluate q_x and q_z as functions of temperature, the quantum theory of specific heats must be used

The heat capacity at constant volume, C_v , for a hexagonal crystal may be considered as being composed of two components, C_{v_x} and C_{v_z} , where

$$C_v = \frac{2}{3}C_{v_x} + \frac{1}{3}C_{v_z}. \quad \dots\dots(4)$$

These two components refer to vibrations perpendicular and parallel to the hexagonal axis, and are expressible by the Debye functions

$$C_{v_x} = 3RD \left(\frac{\Theta_x}{T} \right); \quad C_{v_z} = 3RD \left(\frac{\Theta_z}{T} \right), \quad \dots\dots(5)$$

where Θ_x and Θ_z are the two characteristic temperatures referring to vibrations perpendicular and parallel to the hexagonal axis. Similarly, two components of the heat capacity at constant pressure may be defined by

$$C_p = \frac{2}{3}C_{p_x} + \frac{1}{3}C_{p_z}, \quad \dots\dots(6)$$

the two components again referring to the X and Z vibration directions. Hence, to a high degree of approximation, C_{p_x} is independent of small changes in l_x , and vice versa. We can therefore re-write the integrals (3) as

$$q_x = \frac{2}{3v} \int_0^T \left(\frac{\partial C_{p_x}}{\partial \log l_x} \right)_T \frac{dT}{T}, \quad \dots\dots(7a)$$

$$q_z = \frac{1}{3v} \int_0^T \left(\frac{\partial C_{p_z}}{\partial \log l_z} \right)_T \frac{dT}{T}. \quad \dots\dots(7b)$$

In order to proceed further, C_{p_x} and C_{p_z} must be replaced by expressions involving C_{v_x} and C_{v_z} which, when written as Debye functions, are susceptible to manipulation.

The well-known identity

$$C_p = C_v + \frac{\alpha^2 v T}{\kappa}$$

may be approximated by

$$C_p = C_v + GvT,$$

where G is a constant independent of volume and temperature, since the temperature variations of α , the volume expansion coefficient, and of κ , the volume compressibility, are of the same order of magnitude over the higher temperature range. Hence

$$C_{p_x} = C_{v_x} + G_1 v T, \quad C_{p_z} = C_{v_z} + G_2 v T,$$

where the constants G_1 and G_2 are defined by

$$G = \frac{2}{3} G_1 + \frac{1}{3} G_2.$$

It is now easily shown that, since $v = l_x l_y l_z$,

$$\begin{aligned} \frac{\partial C_{p_x}}{\partial \log l_x} &= \frac{\partial C_{v_x}}{\partial \log l_x} + G_1 v T, \\ \frac{\partial C_{p_z}}{\partial \log l_z} &= \frac{\partial C_{v_z}}{\partial \log l_z} + G_2 v T. \end{aligned}$$

The integrals in the expressions (7) for q_x and q_z therefore become

$$\int_0^T \left(\frac{\partial C_{v_x}}{\partial \log l_x} \right)_T \frac{dT}{T} + G_1 v T \quad \dots\dots (8a)$$

and

$$\int_0^T \left(\frac{\partial C_{v_z}}{\partial \log l_z} \right)_T \frac{dT}{T} + G_2 v T. \quad \dots\dots (8b)$$

To evaluate the C_{v_x} integral in (8a), we express the heat capacity as the Debye function

$$C_{v_x} = 3RD \left(\frac{\Theta_x}{T} \right),$$

whence

$$\begin{aligned} \frac{\partial C_{v_x}}{\partial \log l_x} &= 3RD' \left(\frac{\Theta_x}{T} \right) \cdot \frac{1}{T} \frac{d\Theta_x}{d \log l_x} \\ &= 3RD' \left(\frac{\Theta_x}{T} \right) \frac{\Theta_x}{T} \cdot \frac{d \log \Theta_x}{d \log l_x} \\ &= -3R\gamma_x D' \left(\frac{\Theta_x}{T} \right) \cdot \frac{\Theta_x}{T}, \quad \dots\dots (9) \end{aligned}$$

where $D' \left(\frac{\Theta_x}{T} \right)$ is the derivative of the Debye function with respect to $\frac{\Theta_x}{T}$, and where

$$\gamma_x = - \frac{d \log \Theta_x}{d \log l_x}$$

may be considered as a numerical constant independent of temperature (Grüneisen number). It can now be shown from (9), following a similar procedure to that of Grüneisen and Goens, that

$$\int_0^T \left(\frac{\partial C_{v_x}}{\partial \log l_x} \right)_T \frac{dT}{T} = 3R\gamma_x D \left(\frac{\Theta_x}{T} \right) = \gamma_x C_{v_x} \quad \dots\dots (10a)$$

Similarly,

$$\int_0^T \left(\frac{\partial C_{v_z}}{\partial \log l_z} \right)_T \frac{dT}{T} = \gamma_z C_{v_z} \quad \dots\dots (10b)$$

where

$$\gamma_z = - \frac{d \log \Theta_z}{d \log l_z}$$

From the relationships (7), (8) and (10) we can now express the thermal pressure coefficients in a form suitable for evaluation :

$$q_x = \frac{2}{3} \frac{\gamma_x}{v} C_{v_x} + \frac{2}{3} G_1 T, \quad \dots (11a)$$

$$q_s = \frac{1}{3} \frac{\gamma_s}{v} C_{v_s} + \frac{1}{3} G_2 T. \quad \dots (11b)$$

At low temperatures the terms in T can be neglected and the relationships become simply

$$q_x = \frac{2}{3} \frac{\gamma_x}{v} C_{v_x}; \quad q_s = \frac{1}{3} \frac{\gamma_s}{v} C_{v_s}. \quad \dots (12)$$

The expressions derived by Gruneisen and Goens for low temperatures,

$$q_x = \frac{\bar{\gamma}_x}{v} 3RD \left(\frac{\bar{\Theta}_x}{T} \right); \quad q_s = \frac{\bar{\gamma}_s}{v} 3RD \left(\frac{\bar{\Theta}_s}{T} \right), \quad \dots (13)$$

are essentially similar to the approximate expressions (12). Their definition of $\bar{\gamma}_x$, $\bar{\gamma}_s$, $\bar{\Theta}_x$ and $\bar{\Theta}_s$ is, however, more complex than the simple definition of the similar quantities in this paper, and it is not certain that they have quite the same significance. A comparison of (12) and (13) shows that the γ 's differ by factors of $\frac{2}{3}$ and $\frac{1}{3}$ if the Θ 's are taken the same.

Substitution of the values (11a), (11b) for q_x and q_s in equations (2a) and (2b) gives the final form of the expressions for the two expansion coefficients α_{\perp} and α_{\parallel} . These may conveniently be written as

$$\alpha_{\perp} = \alpha_x = AC_{v_x} + BC_{v_s} + CT, \quad \dots (14a)$$

$$\alpha_{\parallel} = \alpha_s = LC_{v_x} + MC_{v_s} + NT, \quad \dots (14b)$$

where

$$A = \frac{2}{3} \frac{\gamma_x}{v} (s_{11} + s_{12}), \quad \dots (15a)$$

$$B = \frac{1}{3} \frac{\gamma_s}{v} s_{13}, \quad \dots (15b)$$

$$L = \frac{4}{3} \frac{\gamma_x}{v} s_{13}, \quad \dots (15c)$$

$$M = \frac{1}{3} \frac{\gamma_s}{v} s_{33}, \quad \dots (15d)$$

$$C = \frac{1}{3} \{2G_1(s_{11} + s_{12}) + G_2 s_{13}\}, \quad \dots (15e)$$

$$N = \frac{1}{3} \{4G_1 s_{13} + G_2 s_{33}\}. \quad \dots (15f)$$

In computing, A , B , C , L , M and N can be treated as constants.

§3. THERMAL EXPANSION OF GRAPHITE

Magnus (1923) has measured C_p for graphite over the temperature range -229° to 827° C. and has derived the corresponding values for C_v . Using expressions equivalent to equations (4) and (5) of this paper, he found values of Θ_x and Θ_s which gave a good representation of C_p over the whole range. These are

$$\Theta_x = 2280; \quad \Theta_s = 760.$$

These values are used in the following calculations as they are in agreement with our definition and use of the quantities. The experimental data used here have been listed in Part I.

Thermal expansion of the a dimension

The experimental data are such that a reliable plot of α_1 against temperature cannot be obtained. It is preferable, therefore, to consider the a unit-cell dimensions themselves.

The expression (14 *a*) for α_1 will be approximated by

$$\alpha_1 = AC_{v_x} + BC_{v_z}, \quad \dots\dots(16)$$

which assumes that the term directly proportional to the temperature can be neglected. There is some justification for this as the constant C is composed of positive and negative components (s_{13} is negative), both of which are of the same order of magnitude. s_{13} is known to be negative because of the observed shrinkage in the a dimension (see the qualitative discussion in Part I).

Integration of (16) leads to the following expression in a :

$$\log a = AU_x + BU_z + \log a_0, \quad \dots\dots(17)$$

where a_0 is the value of a at absolute zero, and where U_x and U_z , being the integrals of C_{v_x} and C_{v_z} , are components of the internal energy U . U_x and U_z can be evaluated by expressing them in terms of the Debye energy functions which have been tabulated by Beattie (1926) ; thus

$$U_x = 3RT F\left(\frac{\Theta_x}{T}\right); \quad U_z = 3RT F\left(\frac{\Theta_z}{T}\right),$$

where $F\left(\frac{\Theta}{T}\right) = \frac{U}{3RT}$ is the tabulated function.

The problem now is to choose values for A , B and a_0 which give the best

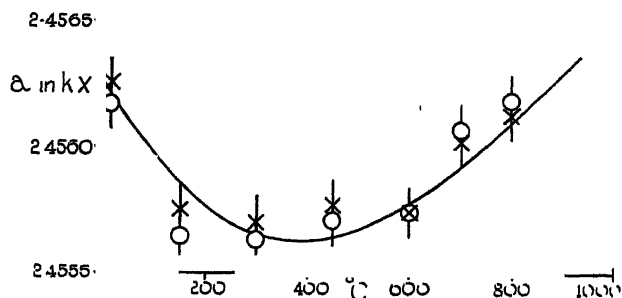


Figure 1. Comparison of observed values of a with the theoretical curve. OX, experimental values from two sets of measurements. Limits of experimental error indicated by vertical lines.

agreement with the experimental data. The curve of a against temperature shown in figure 1 was calculated from (17), using the following values :

$$A = 0.1620 \times 10^{-13}; \quad B = -0.1013 \times 10^{-13}; \quad a_0 = 2.45671 \text{ kx.}$$

The experimental values (two sets of measurements—see Part I), with their

limits of error, are also indicated. The agreement is as good as can be expected considering the limitations of the experimental data.

From the above values for A and B , a plot of α_1 against temperature can be derived from equation (16), using Beattie's tables of $D\left(\frac{\Theta}{T}\right)$ to evaluate C_{v_x} and C_{v_y} . The curve is shown in figure 2. It is in general agreement with experiment, particularly in giving a somewhat greater value for the negative coefficient at 0°C . than for the positive coefficient at 800°C .

Temperature ($^\circ \text{C}$)	α_1 observed	α_1 calculated
0	ca -1.5×10^{-6}	-1.3×10^{-6}
800	ca 0.9×10^{-6}	0.8×10^{-6}

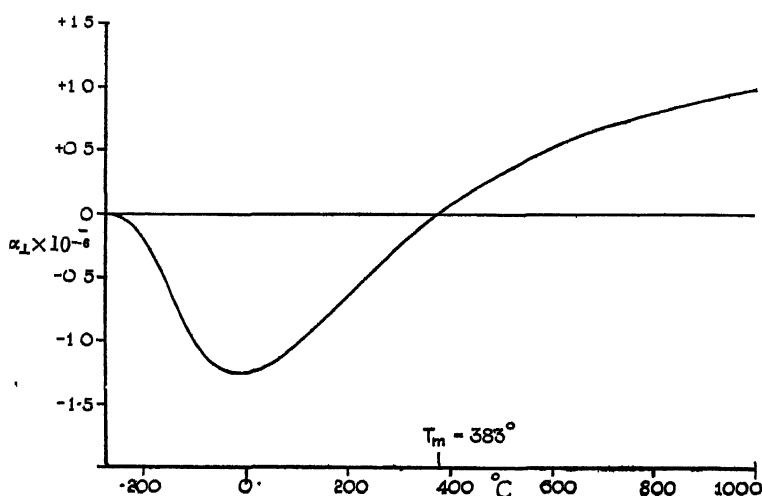


Figure 2. Theoretically derived curve for α_1

The theoretical curve gives $\alpha_1 = 0$ at $T_m = 383^\circ \text{C}$., which agrees with the temperature of about 400°C . found experimentally. It can be said that the theoretical curves agree reasonably well with experiment with regard to the variation with temperature of both a and α_1 .

Thermal expansion of the c dimension

In this case the full expression (14b) for $\alpha_{||}$ was used, which includes a term directly proportional to temperature. The best fit with the experimental data was obtained with the following values for L , M and N :

$$L = -0.770 \times 10^{-13}; \quad M = 1.380 \times 10^{-13}; \quad N = 1.077 \times 10^{-8}.$$

The resulting theoretical curve is compared with the experimental values for $\alpha_{||}$ in figure 3. The agreement is close. Of particular interest is the step-like character of the curve, which was seen in Part I to be demanded by the experimental data. Theoretically, $\alpha_{||}$ must rapidly diminish at temperatures lower than about 400°K ., and must, of course, be zero at 0°K .

As might be expected, it proved impossible to obtain agreement with experiment when the term NT in (14b) was neglected. With such an expression, the

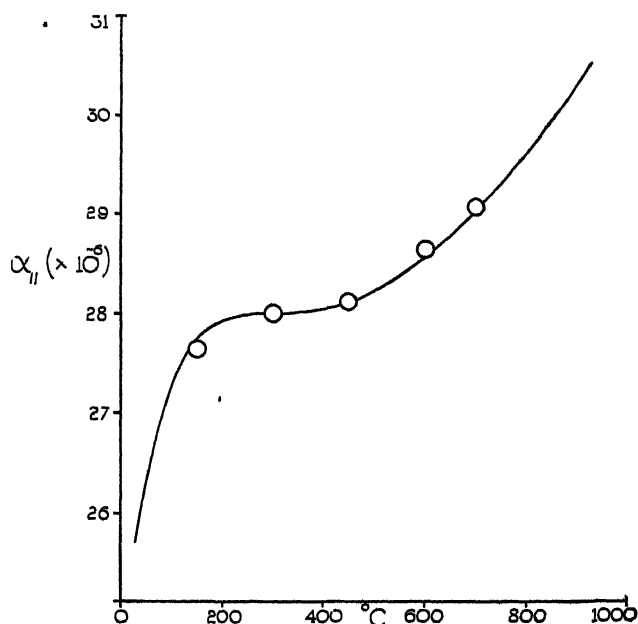


Figure 3. Comparison of experimental values for $\alpha_{||}$ with the theoretical curve.
O, experimental values.

“best” curve for $\alpha_{||}$ rose sharply to a maximum at about 500°K . and then decreased gradually with increasing temperature.

Thermal expansion at very high temperatures

At very high temperatures, the Debye functions approach the value unity. This is expressed more exactly by saying that

$$D\left(\frac{\Theta}{T}\right) \rightarrow 1 \quad \text{when } \frac{\Theta}{T} \rightarrow 0.$$

For graphite, this condition is reached approximately when $T > 2000^\circ \text{K}$. (1700°C). Above this temperature

$$C_v \approx C_{v_x} \approx C_{v_y} \approx 3R.$$

Hence the expansion coefficient can be expressed approximately as a linear function of T :

$$\begin{aligned} \alpha_{||} &\approx (L + M) \cdot 3R + NT \\ &\approx 15 \times 10^{-6} + 1.1 \times 10^{-8} T. \end{aligned} \quad \dots (18)$$

Similarly

$$\begin{aligned} \alpha_{\perp} &\approx (A + B) \cdot 3R + CT \\ &\approx 1.5 \times 10^{-6} + CT. \end{aligned} \quad \dots (19)$$

In this case, if C is negligible, α_{\perp} is approximately constant at very high temperatures, and has a value of about 1.5×10^{-6} .

§4. ELASTIC MODULI AND COMPRESSIBILITIES

The values derived above for A , B , L and M can be used to derive an approximate estimate of the elastic moduli ($s_{11} + s_{12}$), s_{13} and s_{33} , if one additional datum

relating these quantities is known. Such a datum is provided by the volume compressibility. In the hexagonal system

$$\kappa = 2\kappa_{\perp} + \kappa_{\parallel},$$

where κ_{\perp} and κ_{\parallel} are the linear compressibilities perpendicular and parallel to the hexagonal axis. They are related to the elastic moduli by the expressions

$$\kappa_{\perp} = s_{11} + s_{12} + s_{13}; \quad \kappa_{\parallel} = 2s_{13} + s_{33},$$

whence

$$\kappa = 2(s_{11} + s_{12}) + 4s_{13} + s_{33}.$$

In addition, from A , B , L and M we obtain the following quantities by using relations (15 *a*) to (15 *d*) with $v = 5.3$ (volume of a gram-atom of graphite at 18° C.):

$$\gamma_x(s_{11} + s_{12}) = 1.29 \times 10^{-13} \text{ cm}^2/\text{dyne},$$

$$\gamma_x s_{13} = -1.61 \times 10^{-13} \text{ cm}^2/\text{dyne},$$

$$\gamma_x s_{33} = -3.06 \times 10^{-13} \text{ cm}^2/\text{dyne},$$

$$\gamma_x s_{33} = 21.94 \times 10^{-13} \text{ cm}^2/\text{dyne}.$$

Using Basset's (1941) measurement of the instantaneous compressibility at 18° C., $\kappa = 44.9 \times 10^{-13} \text{ cm}^2/\text{dyne}$, we can now derive the following approximate values for the elastic moduli and the two Grüneisen numbers:

$$s_{11} + s_{12} = 1.8 \times 10^{-13} \text{ cm}^2/\text{dyne at } 18^\circ \text{ C.},$$

$$s_{13} = -4.3 \times 10^{-13} \text{ cm}^2/\text{dyne at } 18^\circ \text{ C.},$$

$$s_{33} = 58.5 \times 10^{-13} \text{ cm}^2/\text{dyne at } 18^\circ \text{ C.}$$

$$\gamma_x = 0.71; \quad \gamma_z = 0.38.$$

The two linear compressibilities are, approximately,

$$\kappa_{\perp} = -2.5 \times 10^{-13}; \quad \kappa_{\parallel} = 50 \times 10^{-13} \text{ cm}^2/\text{dyne}$$

There is thus a small negative compressibility perpendicular to the hexagonal axis. The anisotropy of the compressibility is seen to be considerable, the effect of pressure being mainly confined to packing the graphite layers closer together. Values of s_{11} and s_{12} separately are, unfortunately, not obtainable from our data. It is, however, of some interest to have for such an intractable material as graphite even approximate estimates of certain of the elastic moduli.

§ 5 COMPARISON WITH ZINC AND CADMIUM

It is of interest to compare the data for graphite with those obtained for zinc and cadmium by Grüneisen and Goens. Table 1 gives the values of the elastic moduli, compressibilities and thermal expansion coefficients which refer to a temperature of about 18° C.

Table 1

	$s_{11} + s_{12}$ ($\times 10^{-13} \text{ cm}^2/\text{dyne}$)	s_{13}	s_{33}	κ_{\perp} ($\times 10^{-13} \text{ cm}^2/\text{dyne}$)	κ_{\parallel} ($\times 10^{-13} \text{ cm}^2/\text{dyne}$)	α_{\perp} ($\times 10^{-6}$)	α_{\parallel} ($\times 10^{-6}$)
Zinc	7.5	-6.1	28.2	1.4	16.0	13	64
Cadmium	10.8	-9.3	35.5	1.5	16.9	11	100
Graphite	1.8	-4.3	58.5	-2.5	50	-1.2	26

The interesting point emerges that whereas $\kappa_{||}$ is much greater for graphite than for zinc or cadmium, $\alpha_{||}$ is much smaller. This is contrary to the usual state of affairs, as is seen, for instance, by comparing the data for zinc and cadmium. An explanation along the following lines is suggested. Graphite is an extreme example of a layer structure, so much so that it is an assembly of large macro-molecules held together by what are probably mainly van der Waals forces. The covalent linking within a layer is so strong that the layer can be considered to be almost rigid. Thermal vibrations, perpendicular to the layers, then, will be diminished in amplitude owing to the large mass of each layer, and the thermal expansion will not be as great as the extreme anisotropy of the structure would at first suggest. The effect of pressure, however, is independent of the mass of the layer, and a high compressibility results because of the large interlayer spacing (3.35 Å.) and weak interlayer bonding.

The great strength of the C—C bonds within a layer is reflected in the small value of $s_{11} + s_{12}$, whereas the weak interlayer forces give rise to the large value for s_{33} .

The partially covalent nature of the bonds between atoms in the same layer in zinc and cadmium is similarly demonstrated.

Another interesting comparison involves the ratio of interatomic distances within and between layers. In table 2, d_1 is the interatomic distance in Å. within a layer, and d_2 the closest distance of approach of atoms in neighbouring layers.

Table 2

	d_1	d_2	d_2/d_1	$\kappa_{ }/\kappa_{\perp}$
Zinc	2.66	2.91	1.09	11.4
Cadmium	2.97	3.29	1.11	12.1
Graphite	1.42	3.35	2.36	-20

The ratio of the linear compressibilities is obviously correlated with the ratio $\frac{d_2}{d_1}$, while the negative sign of $\frac{\kappa_{||}}{\kappa_{\perp}}$ for graphite is one more indication of the great anisotropy of the structure.

§ 6. ACKNOWLEDGMENTS

The work constitutes part of a research programme of the British Coal Utilisation Research Association carried out under the direction of Dr. D. H. Bangham, who has been unfailing in his help and encouragement. The writer is also indebted to Dr. A. J. C. Wilson for his generous assistance on several points, and to Mr. A. H. Wilson for his critical reading of the manuscript.

RÉFÉRENCES

- BASSET, J., 1941. *C.R. Acad. Sci., Paris*, **213**, 829.
 BEATTIE, J. A., 1926. *J. Math. Phys.* **6**, 1.
 GRUNEISEN, E. and GOENS, E., 1924. *Z. Phys.* **29**, 141.
 MAGNUS, A., 1923. *Ann. Phys., Lpz.*, **70**, 303.
 VOIGT, W., 1910. *Lehrbuch der Kristallphysik* (Leipzig and Berlin. Teubner).

STRUCTURE AND ORIENTATION IN THIN FILMS OF POLYTHENE

By A. CHARLESBY,

Imperial College of Science and Technology, London

MS. received 3 April 1945

§ 1 INTRODUCTION

AN extended series of investigations into the crystal structure and habits of a number of long-chain saturated hydrocarbons has been published, generally limited to paraffins or their derivatives containing from 10 to 50 carbon atoms. While the data obtained in this range constitute an invaluable basis on which to found theoretical studies of intermolecular binding forces, some difficulty arises due to the presence of end groups. A considerable simplification results when the length of each molecule may be considered as infinite, so that each atom bears the same relationship to its neighbours. In common with a number of other properties, the structure of such a hydrocarbon can in a certain measure serve as a limit towards which existing structures approach with increasing length of the $\text{CH}_2 \dots$ chain.

Bunn (1939) has made use of x rays to investigate hydrocarbons of this character and has deduced the electron density of the structure. In the present work, thin films of similar high polymers were studied by means of electron diffraction. In general, good agreement is found in those aspects of the problem which overlap in the two methods of investigation.

A highly polymerized hydrocarbon, polythene, was available for these experiments. It is produced by the polymerization of ethylene, and occurs in a range of molecular weights, which, according to viscosity measurements, amount to several thousand carbon atoms. The commercial product may be obtained in several forms with different average degrees of polymerization. As will be shown later, no appreciable difference was detected between films formed from these different specimens. No attempt was made to obtain specimens whose molecules cover only a very narrow range of molecular weights.

At room temperature the polymer is a white solid, reminiscent of paraffin wax in appearance, but its high degree of polymerisation results in a very tough structure. It shows no definite melting point but softens in the temperature range $90\text{--}100^\circ \text{C}$. At slightly lower temperatures it loses its wax-like appearance and becomes more transparent. As Garner, Bibber and King (1931) have shown, a similar change may be observed in the transition of docosane from the β to α form, which is stable at temperatures several degrees below the melting point. A general formula was deduced by these investigations for the setting point of a hydrocarbon $\text{C}_n\text{H}_{2n+2}$ —

$$T = (0.6085n - 1.75)/(0.001491n + 0.00404).$$

For a molecule of infinite length $T=408^{\circ}\text{K.}$, a temperature somewhat higher than the softening range observed in polythene.

In the present paper, the structure and orientation of polythene at room temperature are discussed. In the following paper the influence of temperature on the structure and the changes which occur in the neighbourhood of the melting point are described.

§ 2 ELECTRON DIFFRACTION—EXAMINATION OF THIN POLYTHENE FILMS

Preparation of specimens

For the purpose of these experiments, polythene was available in two solid forms known as A and B, which differ in their viscosity in solution, and hence in their mean molecular weight. In addition, experiments were carried out on polythene in the form of an emulsion, these will be described later.

In view of the range in molecular weights, and the possible inclusion of grease or other impurities, some degree of purification was judged advisable. Solutions of polythene A (available in the form of rods) were prepared in hot xylene, from which the polythene was precipitated on cooling, in the form of a thick translucent gel. This process was repeated on a number of occasions and specimens from the successive filtrates examined by electron diffraction. In no case was there any appreciable change in either spacing or intensity of the patterns thus obtained, nor did these differ from patterns obtained from specimens of the precipitate. It may therefore be concluded that the polythene under examination does not contain very appreciable quantities of impurity and that it may be considered as a compound sufficiently homogeneous for structure determinations. The inclusion within a pure substance of 1% of a homologous impurity would not, however, have any great influence on the structure.

At first a solution of polythene in benzene was used, but a 0.5% solution of a precipitate from xylene proved far more satisfactory for the preparation of thin films, and this was generally adopted. One or two drops of the solution were placed on hot water (90°C.) and the xylene evaporated off. Films produced in this fashion measured several square centimetres in area, and their thickness, which was variable as judged by their colour, was of the order of 1000 Å. over the major part of the area. The thinnest films produced were of a steely white or golden yellow colour, while thicker films presented a wide and pleasing range of colouring. The films found most useful for transmission were obtained from films of about 1000 Å. thickness. These films are very coherent in spite of their thinness. This may readily be seen when tension is applied to one portion of the film, since the whole film tends to move as a single skin over the surface of the liquid.

When a suitable portion of the film was found, it was removed on a nickel gauze and mounted in the electron-diffraction camera. This was of the Finch type, making use of a cold-cathode discharge tube operating at about 50–60 kv. The distance from specimen to photographic plate is about 50 cm. To produce satisfactory blackening of the plate, exposures of 10–30 sec. were most suitable.

Elucidation of structure

Examples of the electron-diffraction patterns obtained with the film normal to the incident beam are shown in figure 1. The observed spacings are compared with those given by Bunn for a highly polymerized saturated hydrocarbon, as obtained from x-ray examination. The calculated spacings are based on the following axial dimensions :—

Bunn	$a=7.40$,	$b=4.93$,	$c=2.534$.
Charlesby	$a=7.428$,	$b=4.934$,	$c=2.532$.

In examination of these films the three axes may be taken as mutually perpendicular.

One pattern was obtained from a composite film of graphite and polythene, and this was used to standardize the more intense rings of the polythene pattern in terms of the 1.2217 Å. graphite spacing. A number of other patterns were also measured in terms of the 2.467 polythene spacing, which was generally used as sub-standard. The difference from the mean spacings rarely amounted to more than 0.4% and was generally about 0.2%.

Unlike x-ray diffraction patterns of polythene, electron diffraction patterns show sharp rings, there being little or no trace of haloes due to amorphous material. This difference is to be ascribed to the different types of specimens used; films of about 1000 Å. thickness prepared as above are essentially crystalline, whereas polythene prepared from the melt in thick specimens and in the absence of an orientating substratum consists to a considerable extent of amorphous material. The extent of agreement shown for the structure of molecules which may be considered infinite in extent may be seen by comparing them with the corresponding spacings typical of shorter C_nH_{2n+2} chains.

Table 1

n in C_nH_{2n+2}	a	b	
19	7.55	5.01	Müller (1932)
23	7.43	4.97	"
30	7.33	4.92	"
31	7.40	4.93	"
34	7.40	4.95	"
44	7.33	4.93	"
35	7.43	4.97	Hengstenberg
30	7.51	4.955	Schoon
31	7.49	4.97	"

In these measurements a change of 5° C. would cause an error of the order of 1% in the spacings of a and b .

This comparison of the spacings of polythene with those of saturated hydrocarbons shows that the lateral packing of the molecules must be very similar in the two cases. It immediately follows that the polymerized molecule is linear in character.

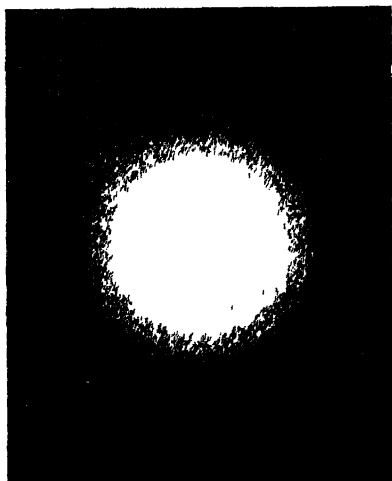


Figure 1.
Polythene film normal to beam

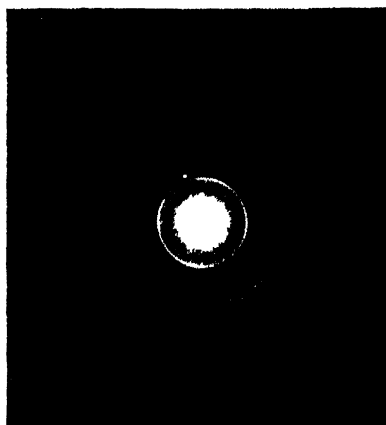


Figure 2
Polythene film inclined to incident beam

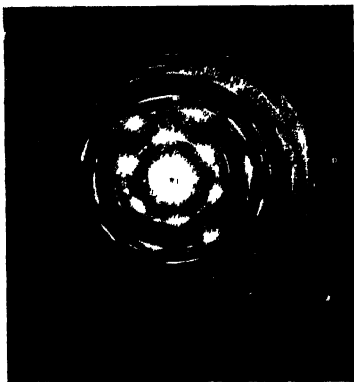


Figure 3. Paraffin wax films grown on water at temperatures of 20° c , 52° c , 100° c

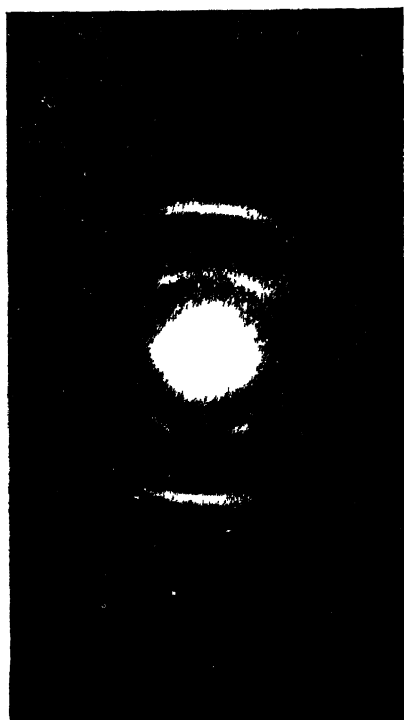


Figure 4
Polythene film fully extended.



Figure 5
Polythene film partially extended.

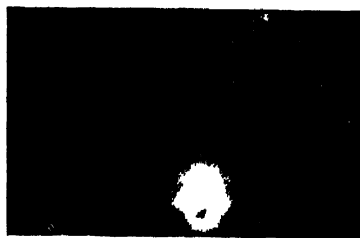


Figure 6. Orientation of polythene on a steel surface
(Incident beam perpendicular and parallel to direction of application.)

This similarity in structure, taken in conjunction with the sharpness of the rings (arising from the relatively large crystal size) and the absence of haloes due to amorphous material, provides strong evidence that during and after polymerization the vast majority of the polythene molecules retain a linear structure without side-chains. The intrusion of a single methyl side-group would of course disrupt the crystal over a wide extent.

While the spacings observed are essentially similar to those of the saturated hydrocarbon studied by Bunn and others, the intensity distribution as between the rings shows considerable differences. For example, among the shorter spacings, the intensity of the 200 is observed to be far weaker and more diffuse than that of the 110 and the 020 diffractions.

Orientation

The above results were deduced with the incident electron beam normal to the polythene film. To investigate the orientation which accounts for the

Table 2. Intensities of rings (visual estimates)

Plates 4006, 4032: Film normal to incident beam

Plates 4008, 4034: Film inclined at about 30°.

Intensities in directions parallel and at right angles to axis of rotation

Plane	4006	4032	4008		4034	
110	VS	VS	VS	S*	VS	VS
200	M	M	M	S	MW	S
210	F	—	—	—	VF	VF
020	S	S	S	—	S	M
120	F	F	—	—	F	—
011	MS	MS	M	M	M	MW
310						
110	M	MW	MW	MS	—	—
201						
220	M	MS	M	M	MS	M
211	MW	MW	MW	MW	W	MW
320						
410	MS	MS	MS	MW	M	M
121						
311	W	W	—	—	W	M

* 110 ring in 4008 (which shows orientation most clearly) is split into four very intense quadrants, at about 45° to the axis of rotation.

The visual estimates of intensity are in the following order.—VS, S, MS, M, MW, W, VW, F, VF.

modified intensity distribution, specimens were tilted through angles of between 20° and 70°. The pattern then showed considerable differences in directions parallel to, and at right angles to, the axis of orientation. These are given in table 2 for certain diffractions of the lowest order. It will be seen that while 020 is weakened if the film is rotated, 200 is intensified. Thus in the film as

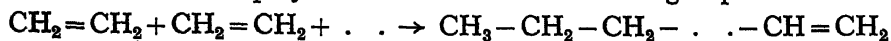
picked up from the water surface, the 020 plane is approximately perpendicular to the surface, but the 200 is inclined, i.e. the b axis lies parallel to the water surface while the a axis is inclined at some angle. This conclusion is verified by an examination of plate 4008, obtained from a specimen the perfection of whose orientation is well above average. Here the 110 diffraction is intensified in directions making angles of about 35° with the axis of rotation. The considerable length of the arcs shows that even in this specimen the orientation is by no means perfect.

Angle of tilt of a axis

Among the large series of paraffins, aliphatic acids and alcohols, the structures of which have been determined, several enantiotropic forms are known to occur. Muller's work on the hydrocarbons has shown that, in a number of cases, transition points occur at certain temperatures, corresponding to a change in structure. Francis, Collins and Piper moreover have noted that in a number of hydrocarbons, the crystallized form depends on the manner in which the crystals were obtained, e.g. on whether crystals were grown from a polar or non-polar solvent. These and similar observations are to be explained as due to the great similarity in the main structure of these hydrocarbons. The main binding force arises from the van der Waals and repulsion forces between neighbouring CH_2 groups. Because of their relative scarcity, side and end groups can only produce considerable modifications in structure under special conditions, e.g. due to steric hindrance. Where this perturbation is not serious, the main chains pack parallel to one another at almost constant distances with the CH_2 groups interleaving. The structure thus obtained is very stable to longitudinal shearing stress, except when its magnitude is such that the CH_2 groups in neighbouring chains ride over one another by an integral number of zig-zags, in which case the binding energy is only slightly altered. In principle, one may therefore anticipate a series of discrete stable or meta-stable structures in each of which the spacings at right angles to the chain length are unaltered.

§3 MOLECULAR ARRANGEMENT OF POLYTHENE

The anomalous intensity distribution observed in the orientation of polythene may be readily explained in terms of the above considerations, if the structure of the polymerized molecule is considered to be substantially linear in character, but that in the course of polymerization two dissimilar end-groups are formed —



While the absence of side-chains permits a lateral packing similar to that prevailing in the saturated hydrocarbons, the unsaturated end-group is too large to be accommodated within the main body of the structure.

In a dilute solution on a water surface the individual molecules orient themselves with the hydrophilic group pointing down into the water. As the solvent evaporates the molecules of the polymer tend to crystallize in the manner customary for a saturated hydrocarbon, but are prevented from so doing by the steric hindrance of the end-groups. This hindrance is, however, avoided in certain monoclinic and triclinic modifications and the molecules are consequently tilted towards the water surface.

In stearic acid films grown on water, two forms α , β are known to occur with the molecules inclined at definite angles to the normal, depending on whether b or a axis remains on the water interface. In the large number of polythene specimens examined, only the α form (our notation $\{nO\}$) has been observed.*

While the end-group may assume a number of different positions, it must move as a rigid unit if there is to be no distortion of bond spacings or angles. The only variation which appears feasible is one of rotation about the penultimate carbon-carbon bond.

Three special cases may be considered :—

- (a) All carbon atoms lie in the main plane of the molecule, and approximately along its axis.
- (b) All carbon atoms lie in the main plane of the molecule, but the C=C bond is approximately at right angles to the axis of the molecule.
- (c) All the atoms of the $-\text{CH}=\text{CH}_2$ group lie in a plane at right angles to the axis of the molecule.

In the saturated hydrocarbon, the closest distance of approach between atoms in different molecules occurs between two hydrogen atoms. The repulsion arising from the interaction between these two atoms plays an outstanding rôle in stabilizing the structure.

In all three cases outlined above, the minimum distance of approach between atoms in adjacent molecules is very much smaller than the value for the saturated hydrocarbon, which is about 2.5 Å. (assuming Bunn's data and a CH bond-length of 1.09 Å.). Even in the most favourable case (a) this distance is only slightly greater than 2.0 Å. To avoid the excessive repulsion arising from this close proximity, while yet retaining the stability inherent in the main hydrocarbon-like structure, adjacent molecules may be sheared to avoid the steric interference. The monoclinic structure which successfully avoids these excessive repulsive forces is the 20. In this structure the angle between the base and the c axis is 56° .

There is no steric hindrance in the b axis direction, and no modification is therefore required. This result is in agreement with the observed structure in which only the angle between the a axis and the water interface is altered from 0° .

That the orientation in polythene is due to the influence of the end-group is rendered even more probable by an examination of the electron pattern obtained from paraffin films produced under similar conditions. While the spacings are similar to those of polythene, the intensity distribution shows very significant differences, due to the $\{0,0\}$ structure of paraffin wax.

Structure of the film

The structure of the film is defined in so far as the side packing of molecules and the orientation at the base are concerned, but the variation in chain length

* Further evidence for the existence of an unsaturated end-group is provided by the discoloration which polythene produces in dilute KMnO_4 .

of the individual molecules precludes the measurement of a crystal height. It appears highly probable that odd fragments or portions of polythene chain protruding from the main structure of each crystal link up with other crystals, forming a dense complex of crystals matted together by molecules common to two or more of them. The general orientation of the chains produced by the basic layer is thus to a considerable extent conserved. Nevertheless, some departure from the strict orientation must be anticipated, departure which becomes more serious in thicker specimens as the influence of the well oriented base layer becomes less effective because more remote. In the thicker films as prepared for x-ray investigation this influence is almost entirely lost in the mass of the solid. For thin films as used in electron diffraction, where the thickness generally corresponds to less than 1000 C—C linkages, the orientation is retained almost in its entirety. Even here, however, the orientation is to some extent impaired as is made clear by the appearance of orders which would be absent in a perfectly oriented crystal.

A point which still remains obscure is related to the length of the individual molecules. In the A polymer the molecule is believed to comprise far more than 1000 C—C groups. It is then difficult to account for the observed structure, without assuming that the molecules are bent back on themselves, and this would presumably disorganize the crystal orientation very considerably. The observed orientation of the film, while by no means perfect, is sufficiently marked to make this hypothesis unlikely.

Influence of the rate of crystallization

Of considerable importance to the structure of the film is the rate and temperature at which crystallization takes place. The films of polythene used in these experiments were prepared on a water surface at 95° c. and formed a skin several square centimetres in size. At lower surface temperatures, when the mobility of the molecules on the water surface is reduced, the film formed is far more brittle, and readily breaks into a number of fragments. Films prepared under these conditions are far less readily mounted and examined.

An analogous effect may be observed in the case of the lower hydrocarbons. A thin film of paraffin wax, prepared from a drop of xylene solution on a water surface at 100° c., is found to consist of relatively large single crystals, measuring about 10^{-2} cm. across, all in approximately the same orientation. On a water surface at 52° c. the crystals formed are several times smaller, while at 20° c. they only measure 10^{-3} to 10^{-4} cm. across. The films of polythene examined all consisted of crystals estimated to measure about 10^{-3} to 10^{-4} cm. across.

Attempts were made to grow polythene crystals by evaporation of the solution on a glass surface. They all resulted in a conglomeration of small milky patches without mutual coherence. Better results might possibly be achieved on a warm glass surface.

Pseudo-hexagonal crystals of paraffin have been observed by crystallization near the melting point, instead of the usual orthorhombic variety, but no such crystals have as yet been observed in polythene films, nor does their existence appear likely.

Samples of B polythene (of lower average molecular weight than A polythene) appear to have a somewhat different texture and appearance. Specimens of B polythene film prepared as above were found to possess the same structure and orientation as A polythene, any slight differences in the patterns, e.g. relative intensities and width of rings, being adequately explained in terms of individual variations due to slight changes in the method of preparation and film thickness.

§ 4. DEFORMATION OF POLYTHENE FILM

When a plastic is deformed, and the stress is subsequently released, there is in general an instantaneous partial recovery, followed by a further slower recovery (after-effect). The final state of the specimen depends on its previous history, the rate and degree of extension, and the conditions under which the applied stress is released. In the study of the elasticity of polymers, temperature plays an exceptionally important rôle.

Polythene exhibits the property of cold-drawing to a remarkable degree. With the electron-diffraction camera it is possible to study the structural changes which take place during the cold-drawing of a thin film of polythene. Owing to the experimental difficulties, the degree of extension cannot be known accurately, but it is nevertheless possible to survey its influence on structure

Thin films prepared as described above were picked up on a small jig consisting of two razor blades mounted in slides in which they could be moved. The initial separation of the knife edges was 0.3 mm. and this was increased during the extension to nearly 1 mm. During this extension part of the specimen was distorted or torn, so that the true extension could not be obtained. A pattern obtained from an extended film is shown in figure 4. In this case the initial colour of the film was yellow and white, the final colour being white.

In the initially unstressed film the molecules make angles of perhaps 60° to 120° with the direction of the subsequently applied force. Electron-diffraction patterns of the extended film show that the molecules assume an orientation nearly parallel to the applied stress, whereas the a and b axes are oriented in all directions at right angles to it. The spacings of the extended and unstretched material were found to be identical within the limits of experimental error.

As shown in figure 4 each hkl reflection consists of an arc. Details of the intensity distribution within each arc are more fully discussed below; it may be stated, however, that the length of each intense arc may be ascribed to imperfect orientation, so that the c axis of each crystal makes on the average an angle θ with the direction of the imposed stress. In one case this angle has been estimated from the length of the arcs. θ is calculated from the ratio $\tan \theta = l/d$, where l is the subtense of the arc, and d is the corresponding diameter. The results are shown below

Table 3

Plane	004	112	002	121	110	200	020
l (mm)	12	5	6	3.5	2	1.5	3
d (mm)	72.4	37.6	36.4	26.7	11.2	12.7	18.7
$\tan \theta$	0.166	0.133	0.166	0.13	0.18	0.12	0.16

Mean value = $9^\circ 1 \pm 1^\circ$

Without a fuller investigation of intensities, it is not possible to state whether the *c* axes of different crystals all make angles of about 9° with the applied stress, or whether, as appears most likely, the angles vary from 0° to 90° according to some form of Maxwellian distribution.

§ 5. CHANGES IN THE STRUCTURE OF THE FILM DURING EXTENSION

The process involved in the extension of thin films of polythene may be deduced from a consideration of several patterns of films extended to varying degrees. In the intermediate stage it is found that 110, 200, 020, etc., diffractions become very weak in directions along the direction of extension, while 002 and corresponding diffractions become intensified. In the perpendicular direction 110, 200 and 020 retain their initial intensities almost unaltered. In intermediate directions the intensities vary in a continuous fashion from 0° to 90° . Thus the re-orientation of each micelle or crystallite takes place smoothly and not by a sudden change to substantially complete orientation.

In films which are almost completely extended, the 100, 200 and 020 are confined to short arcs, but within these the intensities show characteristic changes, which appear to rise from the superposition of short and very intense arcs on longer and weaker arcs. The more intense arcs subtend angles of about 8° ($\pm 4^\circ$ from the mean) with a relatively sharp end, while the weaker arcs subtend about 20° . In the 200 diffraction the weaker arc is extremely faint or missing.

The explanation of these observations is believed to be as follows:—The weaker and more extensive arcs correspond to remnants of the initial rings in the unstretched specimen, and are due to the relatively few micelles or crystals which during the extension remained almost unaltered in position. The stronger arcs arise from micelles whose position has been completely altered, and which finally come to rest in directions almost parallel to that of the extension.

As a result of this motion most micelles lie almost parallel to one another. It is believed that some form of partial recrystallisation takes place in this new orientation. This crystallization cannot be other than very imperfect, in view of the inherent strains and distortions to which the micelles are subject in their new positioning. Nevertheless the presence of very short arcs, which may almost be described as blobs or diffuse spots, and whose ends are relatively sharply defined, does lead one to envisage the existence of intermicellar forces which operate mainly between micelles brought into approximately parallel orientation by the applied stress, and which tend to complete the perfection of orientation when the external stress becomes almost inoperative. It appears very plausible to suppose that the completed blocks, however imperfectly crystallized, are fairly stable, so that the extended film does not revert to its original structure. The intermicellar forces are van der Waals forces which also serve to bind molecules within each micelle. The resultant force must, however, be weaker than the forces within each crystal, due both to the intermolecular strains introduced during the extension, and the greater intermolecular distances generally involved.

When the temperature is raised to a point slightly below the melting point of the pure crystal, these forces no longer suffice to maintain the rigidity of the structure, and there is a resultant tendency for the film to revert at least partially to the initial state.

The reason for the almost complete disappearance of the longer 200 arc is not very clear. A possible explanation is that this may be due to the initial orientation present in the film. The crystals are generally tilted so that the a axis is not parallel to the surface of the film. When the stress is applied, the first tendency is for this tilt to increase. Thus 200 reflections decrease in intensity more rapidly than those due to 020.

In the hypothesis outlined above, it is assumed that each crystal initially present in the film remains substantially unaltered during the extension. Other types of deformation may also be considered. The possibility of a shear within each crystal is rendered unlikely by the observed intensification, as well as by other considerations. An alternative hypothesis, that the extension consists mainly of a rearrangement of amorphous structure within the film is, at least for the thin films studied here, definitely disproved.

B polythene

Extended films of the lower polymer have been obtained but no marked difference in their structure could be detected. From the few specimens prepared, it appears that extended films of B polythene are somewhat less readily obtainable.

§ 6 ORIENTATION OF POLYTHENE

Orientation on metals

The experiments described above are limited to the orientation of polythene on water surfaces and, therefore, provide little evidence for possible differences in orientation produced under different surface conditions. A short series of experiments was therefore undertaken to throw some light on the influence of other substrata and of the treatment to which the film was subjected. Use was made of a metallic base, which offers advantages both from the practical interest in the adhesion of plastic to metals, and for the ease with which several forms of treatment may be applied.

Several experiments were made on a polished copper disc, but in most cases a chromium-plated steel disc was used. The experiments all tend to illustrate the lack of powerful orienting force at the surface, although adhesion is always sufficient to retain some polythene, even when the specimen is subjected to strong rubbing. The results and conclusions are summarized below :—

(a) A chromium-plated steel disc, previously polished, was cleaned by continued washing in benzene. Polythene was then rubbed on the disc in one direction, use being made of a suitably pointed stick of the polymer. The specimen was examined with the electron-diffraction beam parallel to and perpendicular to the direction of rubbing. The two patterns obtained are essentially different in character. With the incident beam parallel, the pattern consists of several rather diffuse rings, whereas with the beam perpendicular it comprises a number of l layer lines, with arcs such as 110, 200, 020, 011, 002, and 022 superimposed. The film is thus found to consist of a series of very small or imperfect crystallites, in which the preponderant orientation is imposed by the molecules which tend to lie along the direction of rubbing, as might indeed be

anticipated on purely mechanical lines. The degree of one-dimensional orientation thus produced is fairly considerable, so that a number of $hk0$, $hk1$, $hk2$ reflections can be clearly distinguished. From the lengths of the arcs involved, it appears that the crystallites make angles of about 10° with the surface (or, alternatively, that the surface is not plane to this extent, although this is far less likely). In addition a very appreciable part of the film consists of less well-oriented crystallites, in which the c axes lie in all directions parallel to the surface.

To within experimental error the spacings involved are identical with those found in films produced by slow evaporation on the water surface, where conditions are far more favourable to crystallisation. On the other hand, the crystals formed are far less perfect or are considerably smaller.

(b) The specimen examined in (a) was subjected to rubbing on cotton wool in the same direction as the polythene had previously been applied. Examination showed that the orientation first found had become even more marked. In particular the rings produced by crystallites in random orientation (but with their c axes parallel to the surface) had disappeared, thus the vast majority of the molecules in the film form crystallites in which the molecular chains lie in directions nearly parallel to each other.

(c) The specimen was then rubbed lightly with cotton wool in a direction perpendicular to the initial direction. The intensity of the pattern in (b) is considerably reduced; in addition intense horizontal layer lines are observed. These are ascribable to a re-orientation of some of the polythene, with the chain standing erect on the substrate. This may either be due to liquefaction of certain regions, with subsequent recrystallisation in a more stable orientation, or possibly to a mechanical re-orientation of the crystals without melting. This latter hypothesis would appear most unlikely.

(d) A further surface as in (a) was prepared and subsequently subjected to hard rubbing with cotton wool in a perpendicular direction. From examination of patterns taken with the electron beam parallel to, and at right angles to, the initial direction of rubbing, it appears that this final application of friction has succeeded in rotating the crystals through $\pi/2$ into the final direction. From the diffuse character of the oriented chains, one may assume the mean length of the crystal (assumed perfect) to be of the order of 100 Å.

These observations can be readily explained in terms of the weak adhesion between polythene film produced in the manner described and the chromium surface. At the same time the re-orientation produced leads one to favour the hypothesis that, by hard rubbing, liquefaction of polythene is produced with subsequently faulty recrystallization in the very limited time available before the temperature falls below a critical value. That this time is very short is shown by the fact that the molecules retain a very clear "memory" of the direction of rubbing, and that this determines their subsequent orientation during the process of crystallization.

(e) A further set of experiments investigated the structure of films produced by evaporation on water, and subsequently picked up on a steel surface. If a

surface film of this nature is intensively dried, and rubbed with wool, most of the film is removed, leaving an invisible layer in which the crystallites are oriented with the c axes in the direction of the applied stress.

(*f*) After heating the thin film of polythene obtained in (*e*) to 180°C ($>$ melting point), the orientation is to some extent destroyed; yet sufficient is retained to indicate quite clearly the initial direction of rubbing. When the film is further heated to 300°C ., no trace of film can be detected.

Orientation of polished copper

Polythene was rubbed on the polished copper disc, and the specimens examined in directions both parallel and perpendicular to the direction of rubbing. The film produced is found to consist mainly of very small crystallites (estimated at about 10 molecules across) oriented in all directions, but with a slight preponderance of molecules standing normal to the surface. The film obtained was generally thin, and this may be related to the poor orientation observed. It must be stressed that the patterns are not as clear as those obtained from the steel specimen, nor is their interpretation so satisfactory. It would, however, appear that adhesion to polished copper is inferior to that of polythene on plated steel.

Orientation on acidified or alkaline water

In view of the orientation always found to occur on water surfaces, orientation which affects the whole crystal structure of the film, the question arises whether this may be changed by altering the $p\text{H}$ of the water.

Films grown on hot acidified or alkaline water (HCl or NH_4OH) do show some weak extra rings, probably due to impurities. The main pattern consists of the usual rings, the spacings and intensities of which show no significant difference from those obtained in the main series of experiments.

§ 7 POLYTHENE ON COLLODION

A short experiment similar to that described above for metals was carried out on a collodion surface. A thin film of collodion prepared by evaporation of a drop of acetone solution on water was picked up on a rocksalt surface, which had previously been polished. On this collodion surface a film of polythene was prepared by evaporation of a drop of xylene solution. The rocksalt was then dissolved away and the composite film mounted on the gauze for examination.

The pattern of the film thus obtained consists of the usual collodion haloes and of the rings of polythene. The intensity distribution of the latter shows that the orientation is similar to that of polythene prepared on a water surface.

In another specimen the polythene film was rubbed prior to the removal of the rocksalt base. Contrary to expectation, no re-orientation of the film was observed. It must, therefore, be concluded either that the use of a collodion substratum prevents the application of considerable pressure, or that gross irregularities in the rocksalt surface protect most of the film.

In a third specimen, solid polythene was directly rubbed on to the collodion backing before the rocksalt base was dissolved away. The pattern of the film was found to correspond to a strongly oriented film in which the molecules lie in the

direction of rubbing, and in which orders up to 004 may be detected. The variation of individual crystals from the main orientation is generally not greater than 5° . In fact the transmission pattern is very similar to that of the stretched specimens of polythene described elsewhere in this paper. Moreover, the sharpness of the rings clearly shows the crystals to be relatively much larger than those formed in a similar fashion on metals.

§ 8 POLYTHENE EMULSION

In addition to the specimens of solid polythene described above, the polymer was also available in the form of an emulsion. This consists of thick white paste, mainly polythene but containing a small percentage of other compounds necessary for stabilization.

The emulsion was diluted in about 100 times its volume of water, and several drops placed on a thin film of collodion or methyl methacrylate as a backing. When the film had dried it was examined in the electron-diffraction camera. In many films the polythene was found to coagulate in small regions, which could not be examined owing to "charging-up" of the specimen in the electron beam. Thinner films suitable for examination gave rise to patterns consisting of diffuse haloes and therefore very dissimilar from the patterns obtained from films of polythene formed by evaporation.

In a few cases some weak rings were also visible but in general these could not be reproduced consistently. These spacings were in approximate agreement with the 210, 020, 120 or 011 and 111, 201, etc., reflections. The intensities of these rings were very low, and moreover other relative intensities differed considerably from those obtained from films produced from a solvent. The intensity and infrequent occurrence of these rings shows that the quantity of crystal polythene present is very small and irregularly distributed. The haloes which form the most striking part of the diffraction pattern were found to correspond to the spacings (after application of Bragg's law) 4.55 (4.2–4.9), 2.2 (2.05–2.35), 1.2 (1.1–1.3) Å. Figures in brackets indicate the approximate width of the halo.

The 4.55 halo is relatively sharp but is superimposed on a weak and broad halo with a peak intensity at about 3.6 Å. At first it was not certain whether the outer haloes were in fact due to polythene or to the collodion or methyl methacrylate background. A comparison with molten polythene, described elsewhere, showed that the 2.2 and 1.2 diffractions are in fact due to polythene, but that the 3.6 halo most probably arises from the backing.

The 1.2 and 2.2 haloes coincide with the centre of a group of rings in the crystalline pattern and may be considered to derive from a blurring of the rings, due to the small size or distortion of the crystals. This explanation does not account for the 4.55 halo, which is clearly distinguishable from the 110 and 200 diffractions.

Electron-diffraction patterns of polythene emulsion show strong similarity to those of liquid polythene. The comparison will be discussed more fully in connection with the latter.

When films prepared as above are heated to 100° C. before examination, no marked changes are found in the electron-diffraction pattern, although faint rings

may be observed somewhat more frequently in the unheated specimens. This change, moreover, was only observed in a few of the heated specimens, and cannot therefore be accepted as a very reliable indication of any increase in crystallinity after heat treatment.

§ 9. NON-ORIENTED POLYTHENE

Attempts were made to obtain films of crystalline polythene which did not show the orientation customary in films produced by evaporation on a water surface. Among the methods attempted were: heating an oriented film, etching in xylene, melting, evaporation on a thin gauze. None of these proved successful in producing a thin film of polythene with an orientation different from that described above.

§ 10. CONCLUSIONS

Thin films of polythene prepared by evaporation from a dilute solution of hot water are crystalline in character. The structure of these films is similar to that of the hydrocarbon described by Bunn, with the following axial spacings in the orthorhombic unit cell: $a = 7\,428\text{ \AA}$, $b = 4\cdot9324\text{ \AA}$, $c = 2\,532\text{ \AA}$. In addition, however, the molecules are oriented by the action of the hydrophilic end-group $-\text{CH}=\text{CH}_2$, which can only be incorporated in the structure if the molecules are tilted, so that the b axis is parallel to the water surface, while the c axis makes an angle of 56° instead of 90° with the water surface. The crystalline character of these thin films of about 1000 \AA thickness contrasts with the amorphous structure of polythene in the bulk form.

Thin films of polythene may be stretched and their altered structure examined. This is found to consist of crystals in which the c axis lies nearly parallel to the direction of the applied stress. The intensity distribution in the electron-diffraction pattern leads one to postulate inter-molecular forces which stabilize the crystals in their new positions, until the temperature is raised to the neighbourhood of the melting point, when the film reverts to some extent to its initial orientation.

On a metal surface, polythene molecules are found to be relatively mobile. They may be oriented mechanically to produce a structure basically similar to that of stretched films.

Polythene emulsion has a structure very different from that of the crystalline film and may, in fact, be described as of an amorphous character. There is, however, some evidence of a small proportion of crystalline polythene. The structure of the emulsion shows great similarity to that of liquid polythene.

ACKNOWLEDGMENTS

The writer wishes to thank Professor G. I. Finch, F.R.S., for advice and Dr. E. H. Rodd and I.C.I. (Dyestuffs) Ltd., for the provision of specimens and a grant during the tenure of which much of the above work was carried out.

REFERENCES

- BUNN, C. W., 1939 *Trans. Faraday Soc.* **35**, 482.
GARNER, W. L., BIBBER, V. and KING, A. H., 1931. *J. Chem. Soc.* 1533.

EFFECT OF TEMPERATURE ON THE STRUCTURE OF HIGHLY POLYMERIZED HYDROCARBONS

By A. CHARLESBY,

Imperial College of Science and Technology, London

MS received 3 April 1945

ABSTRACT An investigation into the effect of temperature on the structure of polyethylene (polythene) is made over a range extending from the temperature of liquid oxygen to temperatures at which polythene exists in the liquid condition. The following main conclusions emerge.

As with much lower polymers, the ratio a/b of the two axes increases with temperature, and tends to the value for a pseudo-hexagonal structure, but this value is never attained, since fusion sets in first.

A change from a crystalline to an amorphous pattern occurs at temperatures which are below those generally taken as the melting point, but which agree with that for which polythene transforms from a wax-like to a more transparent solid. The transition of structure extends over a number of degrees, and in this interval crystalline and amorphous patterns are superimposed.

Orientation present in crystalline films is retained when the films are heated to temperatures well above the melting point, and subsequently cooled.

A preliminary discussion is made on some of the points raised by the observations.

§ 1 INTRODUCTION

IN the previous paper (Charlesby, *Proc. Phys. Soc.* **57**, 496, 1945), the structure of an ethylene polymer of high molecular weight (polythene) was studied at room temperatures by means of electron diffraction. In thin films the polymer presents a structure similar to that of the shorter hydrocarbons C_nH_{2n+2} ; but, due to the enormous lengths of the molecules, the intermolecular distances in the direction of the c axis play an inappreciable rôle in the structure. The existence of the end group $-CH=CH_2$ in the high polymer may, under suitable conditions, modify the structure by tilting the molecules relative to the plane of the film.

The investigation described below is an extension of this work and summarizes the influence of temperature on structure down to the temperature of liquid oxygen and into the region in which polythene exists in the molten condition. Although the results obtained for liquid polythene appear to be of considerable interest, a full critical examination and discussion are postponed to a later paper.

§ 2 APPARATUS

Electron-diffraction patterns were taken with a Finch type of camera. Two modified types of specimen carrier were designed to permit the temperature of the specimen in the evacuated camera to be altered at will. The first type was primarily designed for temperatures below room temperature, but could be used up to about 60 or 80° C. The second enabled the specimen to be heated from room temperatures to well above the melting point of polythene.

The low-temperature design (figure 1) consists essentially of two concentric copper tubes bent upwards at one end and soldered together. The space between these communicates with the main vacuum of the camera at the other end and thus serves partially to insulate the inner tube thermally. The inner tube ends in a heavy brass disc P bearing a screwed rod.

The specimen, usually mounted on a strip of nickel gauze, is clamped to a short holder attached to this rod. It is surrounded by a shield S, which is a close-sliding fit on tube T. In this shield two apertures are provided for the incident electron beam and for the diffracted beams. Each of these apertures is surrounded by guard rings to trap water vapour which might otherwise condense on the specimen at low temperatures.

The device is carried by a brass disc sealed to the outer tubing. It is pressed inwards by the pressure of the atmosphere, and is prevented from sliding downwards by ring R. The carrier C is of the standard pattern; by means of adjustable screws used in conjunction with springs and tombac tubing, it permits the position

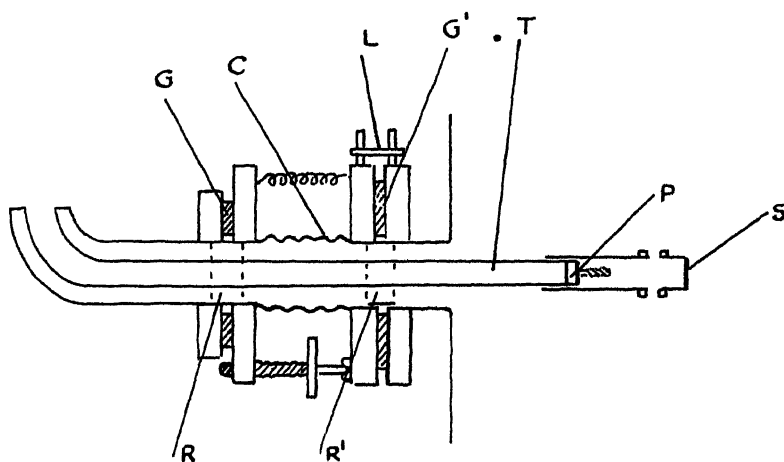


Figure 1

and orientation of the specimen to be adjusted. It is mounted against a standard port of the camera, and is held in position by a ring R' and lug L. Rubber gaskets G, G' are provided to increase the heat insulation of the cooling chamber T.

In operation, the prepared specimen is fixed to the cooling chamber and the shield mounted in a suitable position over it. The complete carrier is fixed in the camera, which is then evacuated. Only when this process is completed is liquid hydrogen introduced into the cooling chamber, as this procedure reduces the possibility of ice condensing on the specimen. The electron beam is scanned over the small portion of the specimen visible through the aperture, and when a suitable adjustment is obtained the plate is exposed. Finally the vacuum is broken and a check is made to ensure that some liquid air remains in the cooling tube, thereby verifying the temperature of the specimen.

For higher temperatures the cooling chamber is filled with hot water through which steam is passed from an external boiler. The temperature is read from a thermometer placed in the water. Unfortunately this method is limited in range

and the temperature readings cannot be relied upon to give the temperature of the specimen, which is fairly distant from the thermometer.

In the construction of the carrier the specimen is situated in what may be considered as a constant-temperature enclosure. Some heat losses occur, mainly by radiation, and wherever possible the apparatus is constructed of heavy metal to ensure good thermal conduction, and hence thermal equilibrium, over the surface. Some very slight temperature differences may arise on the specimen due to direct radiation to or from the camera walls through the apertures. The amount of heat supplied to the specimen by the destruction of kinetic energy of scattered electrons in the incident beam is too small to be appreciable.

Experiments with the first type of apparatus were sufficient to indicate a variation of spacings with rising temperature. The second apparatus permits a more accurate control of temperature as well as a more extensive range. It consists of a demountable furnace, electrically heated in the vacuum, and held at the end of a paxolin rod. As before, the specimen is mounted on a nickel gauze clamped to a short holder, which is screwed to a stout brass plate. The cylindrical furnace also screws on to this plate and is heated by a layer of high-resistance wire wound on asbestos paper. Apertures are provided for incident and emergent electron beams.

A thermocouple passes through the base plate and is fastened to the nickel gauze, which is bent down over it. The temperature within the enclosure is sufficiently uniform, the only appreciable loss being by radiation. Although the thermocouple was mounted several millimetres from the portion of the specimen under examination, there is no reason to believe that the temperature measured was in error by more than a few degrees. The thermocouple was standardized against a standard mercury thermometer both before and after the set of experiments.

§ 3 PROCEDURE

Thin films of A polythene, suitable for electron-diffraction examination, were prepared by evaporation of a drop of xylene solution on a hot water surface. The results are summarized in table 1. Specimens A-G were mounted on nickel gauze, I on a thin methyl methacrylate film mounted on gauze, while H and J were mounted on a very fine meshed gauze.

The change in lattice dimensions of hydrocarbons resulting from thermal expansion has been discussed fully by Muller (1930, 1941) who has shown that any increase in the c axis must be very slight, whilst an increase in both the absolute values of the a and b axes, as well as of the ratio a/b , is to be expected. In the present experiments attention has been confined to change in shape, as, in the absence of a calibration standard, absolute changes in dimensions (which are in any case small) are more difficult to determine than relative changes.

In all the measurements of crystalline polythene given in table 1 the measurements of the 200 (3.714 Å.) and 020 (2.467 Å.) spacings were made relative to 110 (4.109 Å.), chosen as a standard. With increasing temperature the a axis is found to increase in length relative to the b axis, a result in agreement with measurements on short hydrocarbons, and which derives from a tendency to assume a pseudo-hexagonal structure at higher temperatures.

Table 1. Variation of spacing with temperature

Calculated spacings at room temperature														
			4·109	3 714	2 467	2 253	2 055	1 926	1 73	1 61	1 27	1 21	1 098	
Specimen	Plate no.	Temperature (°c)	Observed spacings											Accuracy
			4 109	3·72	2 472	2·249	2·07	1 925	1 73	(1 62)	1 257	1 225	1 10	
A	4921	20	4 109	3·72	2 472				1 73			1 225	1 10	±0·3%
	4922	-183	4 109	3 67	2 485			1·74			1 11			
	4923	-183	4 109	3 73	2 503			1 75	(1 62)		1 11			
D	4929	-183	4 109	3 718	2·49			1·73	1 618			1 24	1 11	±0 5%
	4930	65	4 109	3 745	2·44									
E	4933	72	4 109	3·761	2·472	2·249	2·07	1 925	1·724	1 599	1 257	1 21		±0 3%
	4974	83	4 109	3 78	2·44	2 23	2 06	1 72	1 706	1 598				
	4976	88	4 109	3 74	2 448	2·238								
G	4977	82	4 109	3 774	2 455	2 241	2 06		1 708	1 596	1 251	1 20		±0 6%
	5039	20	4·109	3 71	2 46	2 25	2 07	1·94	1 72	1·605	1·255	1 20	1 10	
	5040	94	4·109	3 78	2 44									
H	5041	111	4·52			2·12					1 24			±0 3%
	5042	107	4 42			2 17					1 21			
	5043	107	4·52			2 11			1 63 ?		1 13			
I	5044	117	(4 6)			(2·1)					(1 15)			-0 5%
	5045	65	4 109	3 7	2 46									
	5059	83	4 109	3·79	2 45									
J	5060	91	4 11		2·44									Strong orientation
		4 6				2 17						1 23		
	5059	107	4·5			2 18						1 23		
	5062	128	4 6			2·17						1·21		
	5063	78	4·5			2·13						1 23		
	5064	20	4 11	3·71	2·46 etc									
						2 21	1·23							
	5065	105	4·7			2 17	1·21							
	5066	169	4·7											
	5067	40	4 11	3 72	2 455 etc									

Remarks.

(1) Calculated spacings at room temperature are based on the following axial spacings: $a=7·428$, $b=4·934$, $c=2·532$ (2) Where possible spacings are measured relative to 110 diffraction ($\pm 109 \text{ \AA}$).

(3) Halo spacings, calculated from Bragg's law, are less accurate.

(4) Specimens A-G were mounted on ordinary gauze, H and J on fine gauze, I on methyl methacrylate film.

(5) In plate 5060 strong haloes co-exist with weak rings showing orientation.

Some estimate of the accuracy of the measurements given in table 1 was obtained by measuring several of the clearer patterns on two or three occasions; the extent of agreement served as a measure of the accuracy involved.

The change in shape is best traced from the change in the ratio a/b , and is relatively greater over the temperature range 20–80° than it is from –183° c.

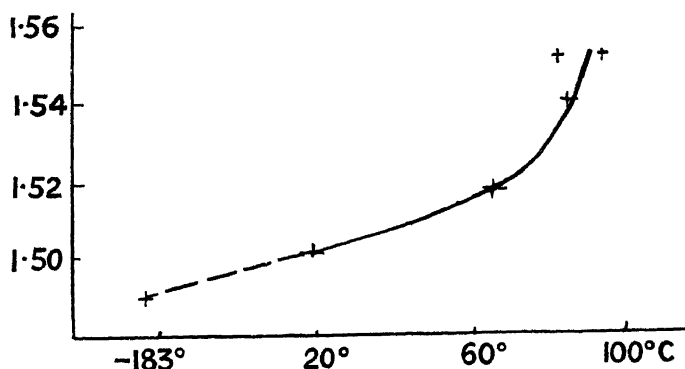


Figure 2 Variation of a/b with temperature

to 200° c. This is to be anticipated from the rapid rise in the repulsive forces between two approaching atoms. As the amplitude of thermal vibrations increases, the skew form of the negative potential-energy curve results in an increased average distance between atoms as shown in figure 2.

§ 4. CHANGES IN STRUCTURE

At temperatures of about 90° c. a profound change in structure occurs, which is evidenced by the transformation from a ring to a halo pattern. The centres of the haloes correspond to spacings of about 4.6, 2.15 and 1.22 Å., as calculated from Bragg's law. These must be identical with the diffuse band at 4.5 to 4.6 observed for liquid hydrocarbons (e.g. Müller, 1932), so that at temperatures well below the point generally accepted as the melting point the structure of polythene must be that of a liquid. It is known that at temperatures of this order polythene assumes a more glass-like appearance, and the change in transparency must be associated with the more amorphous character of the pattern obtained. It is of interest to note that many other transparent polymers, e.g. methyl methacrylate, give rise to electron-diffraction patterns which are very similar in their general appearance, but differ somewhat in the spacings, relative intensities and width of the innermost haloes.

Müller (1932) has shown that as the temperature is raised the structures of hydrocarbons $C_{21}H_{44}$ to $C_{27}H_{56}$ assume a state of pseudo-hexagonal symmetry before reaching the liquid state, but this is no longer true for $C_{29}H_{60}$ and longer molecules, in which the liquid state is first attained. In the hexagonal structure three adjacent molecules define an equilateral triangle ($\phi = 60^\circ$); in polythene the liquid state is reached when ϕ is about 57°.

Before the hexagonal structure is attained, hydrocarbons $C_{24}H_{50}$ and above pass through a transition point, in which the structure shows an abrupt change in dimensions. In $C_{29}H_{60}$, for example, the length of the a axis increases by

5% within a fraction of a degree. The expansion of the polythene lattice has not been traced very closely, but an expansion of this amount does not occur below the melting point. The absence of a transition point for polythene is in any case to be anticipated, since already for $C_{44}H_{80}$ transition and melting points almost coincide.

Of considerable interest is the occurrence together of both halo and ring pattern (e.g. specimen J, plate 5060). This coincidence of the two types of structure may be detected on a number of plates in the neighbourhood of 85° . Over a temperature range of perhaps 10° amorphous and crystalline structure may co-exist in a small portion ($<10^{-2}$ mm²) of a thin film.

This extended range of melting is a strong argument in favour of that theory of the structure of high polymers which envisages the solid as consisting of a number of crystalline regions, held together by molecules common to the structure of several of these regions. Since the forces of cohesion of these molecules must generally be lower than those situated in the bulk of the crystals, the former will first be freed as the temperature approaches the melting point, and form an amorphous or liquid system. The more perfect crystalline regions will continue to exist in equilibrium with the liquid (i.e. amorphous structure) until higher temperatures are reached.

§ 5 COMPARISON WITH HYDROCARBONS OF LIMITED LENGTH

In the absence of absolute measurements of the axial dimensions of polythene at different temperatures, no direct comparison can be made between its expansion and that of shorter hydrocarbons such as $C_{23}H_{48}$ and $C_{29}H_{60}$, studied by Müller. A comparison of the shape of the cross-sectional area may, however, be made in terms of the axial ratio a/b at various temperatures. This is shown in table 2 at room temperatures, at temperatures a few degrees below the melting point, and, in the case of $C_{29}H_{60}$, immediately below and above the transition point.

Table 2. Variation of a/b at various temperatures

	Liquid air or oxygen	Room temperature	Transition point		Below melting point
Polythene	1.49	1 505			1 55
$C_{29}H_{60}$	1.47	1 502	1.513	1 601	1.662
$C_{23}H_{48}$		1.496			1 73

Note · 1 73 corresponds to a hexagonal structure.

It is obvious that the change in dimensions becomes progressively smaller in the higher hydrocarbons, in spite of the more extensive temperature range available.

The comparison is of interest from the point of view of the character of the thermal vibrations of the hydrocarbon chain. In the past it has been suggested that these vibrations occur mainly by a bodily rotation of molecules about their axis. Further work (e.g. Müller, 1940) has shown that tensional vibrations

of the molecular chains must play an important part. In the polythene the length of the molecules precludes the possibility of appreciable bodily rotation. The magnitude of the torsional vibration may to some extent be gauged from the deformation of the structure from its initial dimensions at liquid-air temperatures.

§6 PERSISTENCE IN ORIENTATION

All the films studied were produced by evaporation on the water surface and showed the {20} form of orientation produced under these circumstances. This orientation is readily detectable by a study of the relative intensities of the diffraction rings, especially if the film is not normal to the incident electron beam. Present experiments permit an investigation into the extent to which this orientation is conserved at higher temperatures.

In tilted specimens the initial orientation present may be best judged from the arcing of the 020 reflection, which is most intense in a direction parallel to the axis of rotation of the specimen. In all halo patterns no orientation is detectable, even in tilted specimens.

At low temperatures (-183°C.) the orientation pre-existing at room temperature remains unaltered. As the temperature is raised to 80°C. , there is no appreciable change in the orientation, the 020 remaining arced, although there is an overall reduction in the intensity of the reflections relative to the 110.

In specimen I an examination of the structure was made after the specimen had remained in the amorphous condition (corresponding to a liquid structure) for a period of nearly one hour. Some orientation was found to persist on cooling to 65° , even though the specimen had previously reached a temperature of 117° . In this specimen the polythene film was spread on the surface of methyl methacrylate, and the presence of the latter might conceivably cause a re-orientation on cooling. Specimen J was therefore prepared on a support consisting of a fine nickel gauze, and was heated to 128°C. The orientation of this film after cooling was most marked. Finally a temperature of 168° was reached, but even after this treatment the remaining orientation was still strong.

It may be suggested that re-orientation occurs subsequently to cooling under the influence of the gauze support or of surface tension. Neither of these suggestions can, however, account for the type of tilted orientation actually observed, which is typical of films oriented under the action of the water surface. In a globule of molten polythene in which the molecules are arranged at random no forces can exist which, when the film is cooled, re-orient the molecules at an angle to the surface similar to that produced on a water surface.

The important conclusion is therefore reached that in thin polythene films, and probably in all molten high polymers, the initial structure of the solid is to a considerable extent preserved in the molten form for long periods. The oriented molecules may then serve as a nucleus of condensation in the subsequent process of solidification.

The present experiments give strong evidence in favour of the hypothesis that crystal structure is to some extent retained in the liquid form.

§7. STRUCTURE OF LIQUID POLYTHENE

X-ray diffraction patterns of hydrocarbons consist of a diffuse halo, the spacing of which corresponds to 4.5 to 4.6 Å. if Bragg's law is used to translate distances measured on the x-ray plate into spacing between planes. The haloes obtained from polythene films at temperatures of 90° C., and which persist at least up to 168° C., must be of a similar origin. In addition to the 4.6 Å. halo, haloes corresponding to spacings 2.15 and 1.22 Å. are observed in the electron-diffraction patterns. This increase in the data provided renders a search for a suitable structure more hopeful.

The 2.15 Å. halo occurs at a spacing overlapping reflections such as 111, 201, etc., and its presence may be explained as deriving from these reflections in a distorted or minute structure. Moreover, the 1.22 halo almost certainly corresponds to the 002 reflection, with, perhaps, an overlap from such reflections as 202, etc. The origin of the 4.6 Å. halo nevertheless remains obscure.

The internal structure of liquid hydrocarbons is far from clear. If it partakes of a crystalline character, the minute size of the crystals accounts for the diffuse character of the reflections observed. The application of Bragg's law is then at least approximately justifiable, but the position of the haloes cannot be readily explained. On the other hand, if the structure is essentially amorphous in character, the application of Bragg's law requires further justification.

§8. POLYTHENE IN THE FORM OF A LIQUID AND EMULSION

The pattern of polythene emulsion at room temperature shows a remarkable similarity to those of molten films of polythene, as is shown in table 3.

This great similarity indicates that the intermolecular spacings and orientation must be similar, and hence that polythene emulsion exists as a supercooled form of liquid polythene.

A further specimen of polythene containing a very small percentage of extraneous matter was examined after standing for about one year. The patterns obtained were more nearly crystalline with spacings corresponding to those of crystalline films. The sharpness of the rings showed that the crystals could not be very extensive. Moreover, the intensity distribution was such that the orientation of the films produced by evaporation on a water surface no longer existed. In most cases only the 110 reflection was observed.

Table 3

		Å	Å	
Molten polythene	Average of	4.6	2.15	1.22 Centre of haloes
	10 plates	4.1-5.0	2.05-2.25	1.10 Width of haloes 1.35
Emulsion	Average of	4.6	2.2	1.2 Centre of haloes
	2 plates	(3.6)		

§9 CONCLUSION

As the temperature of a film of polythene is raised, its structure is modified by the tendency to form a pseudo-hexagonal structure. This is not, however, achieved, as melting sets in at a lower temperature.

Polythene in the molten condition presents an amorphous structure, represented in the electron-diffraction pattern by three haloes at 4.6, 2.15 and 1.22 Å., that at 4.6 Å. being the most sharp and intense. Over a range of several degrees the amorphous and crystalline types of structure co-exist throughout the film. That the melting point of polythene is not sharp is common knowledge, but it is of interest to find this associated with a gradual transition from the fully crystalline to a fully amorphous structure. In shorter hydrocarbons the transition is far more sudden. It may be argued that molecules of varying lengths form corresponding crystals with varying melting points, but this cannot be supported by the evidence, since the temperature region in which the transition from crystalline to amorphous structure occurs is far lower than that to be anticipated for crystals of lengths approximating to those of the highly polymerized molecules. If the change in pattern is due to breakdown of crystal structure, the individual crystals must be far shorter than the individual molecule. Thus, at room temperature, molecules must enter into the structure of a number of these crystals.

It is then reasonable to assume that some molecules are more loosely held, and on being released from many of their bonds at temperatures below the melting point of the bulk crystal form the amorphous phase. As the temperature is further raised, the proportion of crystalline structure rapidly decreases.

The region in which the transition from crystalline to amorphous structure occurs extends over a number of degrees, but, certainly as far as can be judged from the electron-diffraction pattern, the change is complete at temperatures below the melting point as the latter is generally understood, and, moreover, appears to coincide with the usual transition to a more glass-like material.

The conservation of the orientation initially present in the films even when these are heated to 168° C. shows that the molecular arrangement is not entirely random even at temperatures well above the melting point. Further work is clearly desirable on the rate at which the initial orientation is lost when the film is held in the molten state for varying periods. These considerations lead one to envisage a statistical fluctuation of temporary crystallinity over small regions, sufficient to ensure mobility above the melting point, over macroscopic periods of time, yet serving to retain some "memory" of initial structure.

REFERENCES

- CHARLESBY, A., 1945. *Proc. Phys. Soc.* 57, 496
MULLER, A., 1930 *Proc. Roy. Soc. A*, 127, 417, 1932. *Ibid.* 138, 514; 1940.
Ibid. 174, 137 1941, *Ibid.* 178, 232

NOTE ON DIFFUSION IN THE IONOSPHERE

By J. C. JAEGER,
The University of Tasmania

MS received 31 May 1945

ABSTRACT The equation of diffusion for a one-dimensional region in which the coefficient of diffusion varies exponentially is solved, and the results applied to the diffusion of ions in the ionosphere. Numerical results are given for three special initial distributions of ion density.

THE effect of diffusion of ions in the ionosphere has been very little studied. The general equations, including diffusion, recombination and ion production, were set up by Hulburt (1928) but are quite insoluble. Recently Bagge (1943) has shown, by numerical integration in a special case, that diffusion will greatly reduce the ion density in the upper parts of a Chapman region, although it leaves that near the maximum unaltered.

The object of this note is to set out briefly the results for one case in which the exact solution is possible, namely, that of an atmosphere in which the particle density varies exponentially with height, and in which the effects of recombination and the earth's magnetic field are neglected. This allows an estimate of the effect of diffusion to be made which may be useful in the higher regions, where recombination is of less importance. The neglect of the earth's magnetic field is, of course, serious; its effect is considered briefly later.

Let N be the number of ions per c.c. at height z , and let $1/b$, supposed constant, be the scale height of the atmosphere, so that the coefficient of diffusion at height z , which is inversely proportional to the particle density at this height, may be written

$$\kappa e^{bz}, \quad \dots\dots(1)$$

where κ is a constant. Then if $I(z, t)$ is the rate of ion production at height z at time t , the differential equation satisfied by N is [cf. Hulburt (1929) or Bagge (1943)]

$$\kappa e^{bz} \left(\frac{\partial^2 N}{\partial z^2} + 2b \frac{\partial N}{\partial z} + b^2 N \right) - \frac{\partial N}{\partial t} + I(z, t) = 0. \quad \dots\dots(2)$$

Since in practice we are only interested in large values of z , and those ions which diffuse to smaller values of z will be removed by recombination, it is sufficient to consider (2) in the region $-\infty < z < \infty$ with the boundary conditions N finite as $z \rightarrow \pm \infty$. We proceed to solve (2) in this region for $t > 0$, with a given initial value $N_0(z)$ of N when $t = 0$, for the case* $I(z, t) = 0$.

* There is no additional difficulty in including ion production, but it is perhaps pushing a simple model too far.

Writing

$$y = Ne^{bz}, \quad x = e^{-bz}, \quad \dots\dots (3)$$

(2) becomes in the case $I(z, t) = 0$,

$$\frac{\partial}{\partial x} \left(x \frac{\partial y}{\partial x} \right) - \frac{1}{\kappa b^2} \frac{\partial y}{\partial t} = 0, \quad 0 < x < \infty, \quad t > 0, \quad \dots\dots (4)$$

to be solved with

$$y \text{ finite as } x \rightarrow 0, \text{ and as } x \rightarrow \infty, \quad \dots\dots (5)$$

and with

$$y \rightarrow y_0(x) = \frac{1}{x} N_0 \left(\frac{1}{b} \ln \frac{1}{x} \right), \quad \text{as } t \rightarrow 0. \quad \dots\dots (6)$$

The solution* of (4), (5) and (6) is

$$y = \frac{1}{\kappa b^2 t} \int_0^\infty y_0(u) I_0 \left(\frac{2\sqrt{(ux)}}{\kappa b^2 t} \right) e^{-(x+u)/\kappa b^2 t} du. \quad \dots\dots (7)$$

We shall write k for the dimensionless quantity:

$$k = \kappa b^2 t. \quad \dots\dots (8)$$

It is easy to verify that (7) satisfies the differential equation and initial and boundary conditions.

To illustrate the behaviour of the solutions of (4) we consider two cases in which (6) takes a specially simple form.

Case (i). $N_0(z) = n_0 \delta(z - z_0)$, where δ is the Dirac δ function, that is, a concentration of n_0 ions is released at $t = 0$ at the level z_0 . In this case the solution is

$$\frac{N}{bn_0} = \frac{x}{k} I_0 \left(\frac{2\sqrt{(xx_0)}}{k} \right) e^{-(x+z_0)/k}, \quad \dots\dots (9)$$

where $x_0 = e^{-bz_0}$

The right-hand side of (9) as a function of x involves only the quantity k , but if we express it in terms of z , b appears as well. As the results are a little more physical when expressed in terms of z , they are plotted in this way in figures 1 and 2 for the value $b = 2 \times 10^{-7}$ corresponding to a scale height of 50 km., that usually observed in the F region. Figure 1 shows the diffusion of a concentration of n_0 ions initially at $z_0 = 200$ km., and figure 2 that of a concentration initially at $z_0 = 400$ km., for various values of $k = \kappa b^2 t$. A large downward motion is at once apparent. Whether this is of physical importance depends on whether values of k of the order of those for which the curves are drawn can be attained in practice. The coefficient of diffusion at any height is $0.41\gamma u$, where γ is the mean free path at that height and u the mean molecular velocity. Taking the value of γ at 250 km. to be 10^6 cm. (Hulburt (1937), table 1) and for u its value at 220° K., 10^5 cm. per sec., we get 4.1×10^{10} for the value of the coefficient of diffusion.

* For the method of solution using the Laplace transformation, see Carslaw and Jaeger, *Operational Methods in Applied Mathematics*, § 68 (Oxford, 1941) $I_0(x)$ is the modified Bessel function of the first kind of order zero.

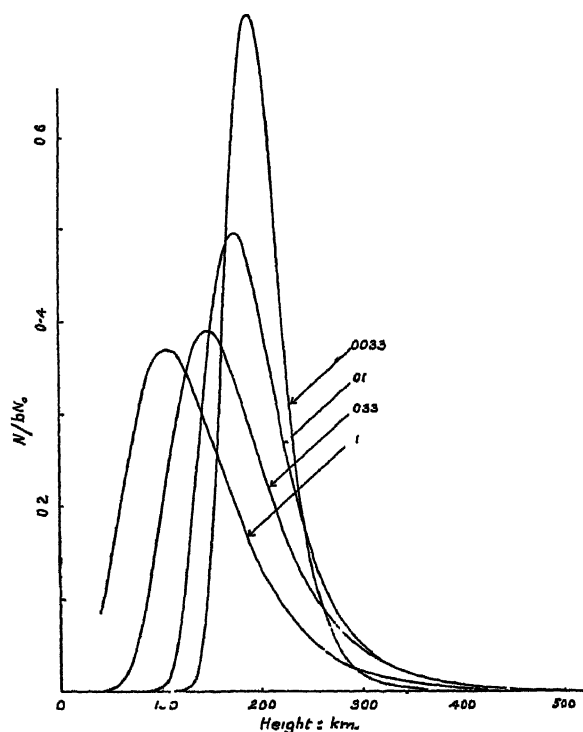


Figure 1 Diffusion of a concentration of ions initially at 200 km. Scale height 50 km.
The numbers on the curves are the values of k

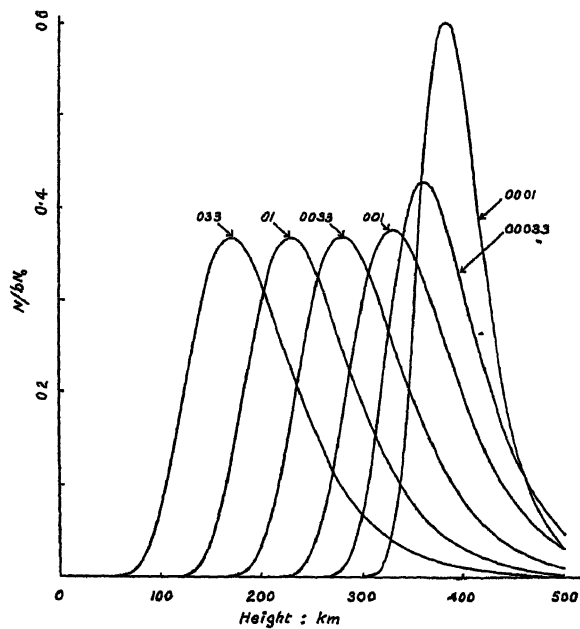


Figure 2. Diffusion of a concentration of ions initially at 400 km. Scale height 50 km.
The numbers on the curves are the values of k

If we give b the value* 2×10^{-7} obtained from the scale height, 50 km., of the F region, we have from (1), for $z = 2.5 \times 10^7$,

$$\kappa e^5 = 4.1 \times 10^{10},$$

that is, $\kappa = 2.8 \times 10^8$. It follows that, with these values of κ and b , k has the value 0.04 at the end of an hour, so that values of k of the order of those in figures 1 to 3 seem quite possible.

Case (ii). Another case in which the integral (7) can be evaluated simply is that in which the initial distribution is of the familiar Chapman type, say

$$N_0(x) = N' x^{\frac{1}{2}} e^{-cx}, \quad \dots (10)$$

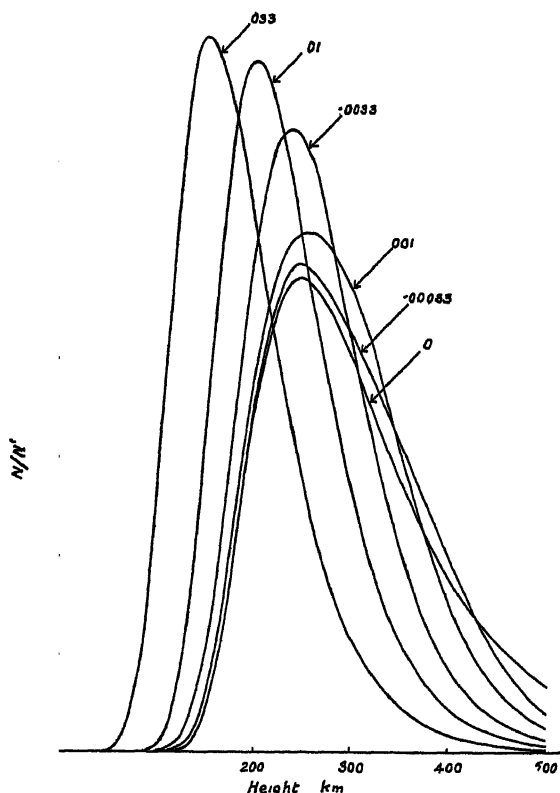


Figure 3. Diffusion of a distribution of ions which is initially of Chapman type. The numbers on the curves are the values of k .

where x is given by (3). In this case we find

$$N = N' x (\pi/kk')^{\frac{1}{2}} e^{-x(2k'-1)/2kk'} I_0(x/2kk'), \quad \dots (11)$$

where k is defined in (8) and $k' = 1 + kc$.

In figure 3 curves of (11) for the case $c = 80$, $b = 2 \times 10^{-7}$, and various values of k are given. The tendency for the upper ions of a theoretical Chapman region to move downwards by diffusion as described in Bagge (1943) is clearly shown.

* If values of b based on the kinetic theory are used, for example those deduced from Hulburt's (1937) table 1, much smaller values of k are obtained, but these do not allow for the probable high temperature at these levels.

The effect of the earth's magnetic field is to leave the coefficient of diffusion unaltered for diffusion along the magnetic field, but, for diffusion across the magnetic field, in place of κe^{bz} , we have

$$\frac{\kappa e^{bz}}{1 + ae^{2bz}},$$

where a is a known constant depending on the field. The resulting equation is not soluble in terms of known functions. In temperate latitudes the effects described above will not be altered greatly, but at the magnetic equator vertical diffusion will not be an important agent for removing ions from high levels

REFERENCES

- BAGGE, E., 1943 *Phys. Z* **44**, 163
 HULBURT, E. O., 1928 *Phys. Rev* (2) **31**, 1018
 HULBURT, E. O., 1937 *Rev. Mod. Phys.* **9**, 44

THE PHOTOGRAPHIC ACTION OF X RAYS

By S. R. PELC,

Physics Department, Hammersmith Hospital, London

Communicated by Dr L. H. Clark, 20 May 1945

ABSTRACT. An account of theoretical considerations of the photographic action of x rays, based on the absorption process of x rays in photographic emulsions, and found to be in good agreement with experimental results, is presented in this paper. An equation giving the variation of the photographic action with wave-length is deduced. Consideration is given to the number of grains made developable per absorbed photon. The conclusion is reached that the energy required to make one grain developable is, for a given emulsion, constant for effective wave-length from 47 x u. to 395 x u

§ 1. INTRODUCTION

THE results of previous investigations show that the shape of the density-exposure curve is essentially the same for different qualities of radiation (Glocker and Traub, 1921). Berthold (1925) found that in order to produce the same density with x-ray beams of different penetrating power, more energy must be absorbed in the emulsion from a penetrating radiation than from a less penetrating one. Similar investigations (G. E. Bell, 1936; E. E. Smith, 1943) with heterogeneous x rays show that the exposure, expressed in roentgens, required to produce a certain density passes through a minimum in the region of 0.25 mm. copper H.V.L.

M. Blau and K. Altenburger (1922) and L. Silberstein (1922) assumed that at any stage of the exposure the number of affected grains (n) is proportional to the number of unaffected grains, i. e.

$$dn = (N - n)\epsilon dI, \quad \dots\dots(1a)$$

where N is the total number of grains per unit area, ϵ is an arbitrary factor and I the exposure.

On integration, we find

$$n = N(1 - e^{-\epsilon I}). \quad \dots\dots (1b)$$

L. Silberstein (1922) regarded each grain as presenting to the photons a target of area a , and gave equation (1b) in the form

$$n = N(1 - e^{-a\beta Q}) \quad \dots\dots (1c)$$

where Q is the number of incident photons and β an arbitrary factor. The meaning of β has been differently interpreted: by L. Silberstein and A. P. H. Trivelli as the fraction of the area of a grain which acts as an efficient target for light quanta, and by A. Charlesby as the probability of a grain becoming developable when struck by a photon. The density is taken as proportional to the number of grains, and it was found that the density-exposure curves for x-ray exposures can be well represented for moderate densities by an equation of the form of equation (1b) (Bell, 1936, Charlesby, 1940). It can be shown that the speed of a photographic emulsion determined in the usual way from the curve of density against log (exposure) is proportional to ϵ whenever the density-exposure curve can be represented by (1b), ϵ is sometimes referred to as the *speed* in this paper. The following considerations show that the arbitrary factor ϵ can be evaluated from the absorption process of x rays in the silver halide of the emulsion and the known number of incident photons per unit area, when the exposure is expressed in international roentgens.

§ 2 THEORETICAL

The number of grains affected, n , will be proportional to the number of quanta absorbed, and to the number of grains which each absorbed quantum renders developable, as well as to the probability of a hitherto unaffected grain being affected. Taking q to be the number of photons absorbed and η the number of grains affected per quantum absorbed, then

$$dn = \frac{N-n}{N} \eta dq. \quad \dots\dots (2)$$

From the known absorption of x rays ($I = I_0 e^{-\mu d}$), the number of quanta absorbed is found to be

$$dq = (1 - e^{-\frac{\mu}{\rho} g}) dQ, \quad \dots\dots (3)$$

where Q is the number of incident quanta, μ/ρ the mass-absorption coefficient of silver bromide or silver chloride, and g the total amount of silver halide in grams per sq. cm. of emulsion. The amount of silver in the emulsion is of the order of 10^{-4} per sq. cm., so that for the qualities of x rays used in this investigation, (3) can be simplified to (3a) with an error not exceeding 1%.

$$dq = \frac{\mu}{\rho} g dQ. \quad \dots\dots (3a)$$

The total amount of silver bromide is $N\gamma$, where γ is the mass of the undeveloped grain, and substituting (3a) in (2) gives

$$dn = (N-n) \eta \frac{\mu}{\rho} \gamma dQ. \quad \dots\dots (4)$$

For x rays measured in international roentgens (r.) the number of quanta per sq. cm. cross-section of the beam can be calculated from the energy necessary

to liberate one pair of ions in air (35.5 e.v.) and from the definition of the roentgen, and is found to be

$$Q = \frac{\lambda r}{(\tau + \sigma_a)_{\text{air}}} \times 0.597 \times 10^7, \quad \dots (5)$$

where λ is the wave-length in Angstrom units, τ the photoelectric absorption, and σ_a Klein-Nishina's scatter-absorption coefficient in air.

Substituting the value for the number of incident quanta (5) in equation (4) gives the final expression,

$$dn = (N - n)A\eta\gamma dr, \quad \dots (6)$$

where

$$A = \frac{\mu_{\text{AgBr}}\lambda}{(\tau + \sigma_a)_{\text{air}}} \times 0.923 \times 10^6. \quad \dots (6a)$$

Integration from 0 to r yields

$$n = N(1 - e^{-A\eta\gamma r}). \quad \dots (7)$$

Comparing equation (7) with equation (1b) we see that the exponential factor ϵ is now expressed as a function of separately assessable terms and elementary constants when the exposure is expressed in international roentgens.

$$\epsilon = A\eta\gamma. \quad \dots (7a)$$

The proportionality between the speed (ϵ) and the average mass per grain as indicated in equation (7a) is obviously in contradiction to the work of L. Silberstein and A. P. H. Trivelpiece (1930) and of A. Charlesby (1940), and will be referred to in more detail in the discussion of results.

For photographic work, the measurement of densities is, of course, preferable to counting the number of grains, and a few remarks about the connection between the number of affected grains, development and the resulting density may be inserted here. In general only a fraction α of the affected grains will be developed, and equation (7) can be written

$$n_d = \alpha n = \alpha N(1 - e^{-\epsilon r}). \quad \dots (7b)$$

The fraction α can be calculated from the increase of density with time of development.

From a consideration of the combined transparency of a number of layers in a photographic emulsion, P. G. Nutting (1913) deduced the formula

$$d = -p \log \left(1 - \frac{S}{p} \right), \quad \dots (8)$$

where d is the observed density, p the number of layers, defined as the thickness of the emulsion divided by the average thickness per grain, and S the total projected area of the developed grains. If S/p is small compared with p , equation (8) can be approximated by Nutting's simplified formula,

$$d = 0.4343S. \quad \dots (9)$$

The number of grains affected, as expressed in equation (7), and the corresponding density as expressed by (8) in connection with (7), will be strictly valid for emulsions consisting of grains of equal size and approximately valid for emulsions with relatively small variation in grain-size.

For measurements in the region of low densities, when the density can be well approximated by equation (9), equations (7*b*) and (9) combined give

$$d = 0.4343\alpha\bar{a}N(1 - e^{-g}), \quad \dots\dots(10)$$

where \bar{a} is the average projected area per grain. For small values of the term in the exponent, and low densities, i. e. as far as the familiar linear portion of the density-exposure curve is in the region of low densities, equation (10) can be further simplified to

$$d = 0.4343\alpha\bar{a}N\Delta\eta\gamma r. \quad \dots\dots(11)$$

Substituting $g = N\gamma$ as before gives the most convenient expression for low densities in slow emulsions of high contrast, namely,

$$\frac{d}{r} = 0.4343\alpha\bar{a}\Delta\eta g. \quad \dots\dots(11a)$$

The error incurred in these approximations calculated for the emulsions used in this investigation from the known difference between equations (8) and (9) and between equations (10) and (11) respectively is up to 4% for one of the emulsions chosen for experimental investigation (film Y) and below 1% for film X for densities smaller than 0.5; the simplified equations were therefore used for film X only.

§ 3 EXPERIMENTAL

The equations resulting from the theoretical considerations of the first part contain, with the exception of the quantum efficiency η , factors which can be either calculated or determined without recourse to x-ray exposures. To test the validity of the theory an experimental investigation into the dependence of the photographic action of x rays on wave-length, and a comparison of measured densities with calculated values was undertaken.

Two types of commercial film with widely different characteristics, as shown in table 1, were chosen for experimental investigation.* The films were exposed to x rays in a special cassette made from $1/16$ -inch brass sheet, which permitted

Table 1

Film	Total amount of silver bromide (gm per sq cm.)	Average diameter of the developed grains (cm.)
X	6.2×10^{-4}	0.8×10^{-4}
Y	2.6×10^{-4}	3.0×10^{-4}

the exposure of a strip of 2 cm. and accommodated up to 25 mm. thickness of lead above the part of the film not intended for exposure. The film was enclosed in thin black paper, backed by about $1/16$ inch of cardboard and slipped into a black envelope, similar to the usual double wrapping. No influence from back-scatter from the backing materials could be observed. The exposures were controlled by ionometric measurements, for which instruments calibrated at

* Thanks are due to Mr R. Tupper, of the Department of Chemistry, British Postgraduate Medical School, who determined the amounts of silver per unit area quoted in tables 1 and 3.

the N.P.L. against the standard free-air chamber were used. The error of the ionometric measurements was $\pm 3\%$ for radiations up to H.V.L. = 2 mm. Cu, and is unlikely to exceed $\pm 5\%$ for the harder radiations. The radiation from three x-ray tubes, excited by constant-potential generators, covering a range from 90 to 500 kv., were used in order to obtain x-ray beams within a wide range of penetrating powers for the exposures; the target-film distances were varied between 60 and 150 cm. to allow suitable times of exposure. In some of the series, standard tints exposed on each film served to eliminate the possible influence of differences in development, and in all cases all films belonging to one series of experiments were developed in the same developer within as short a time as possible. Specular densities measured with a microphotometer and barrier-layer cell calibrated against a step-wedge could be reproduced within $\pm 2\%$.

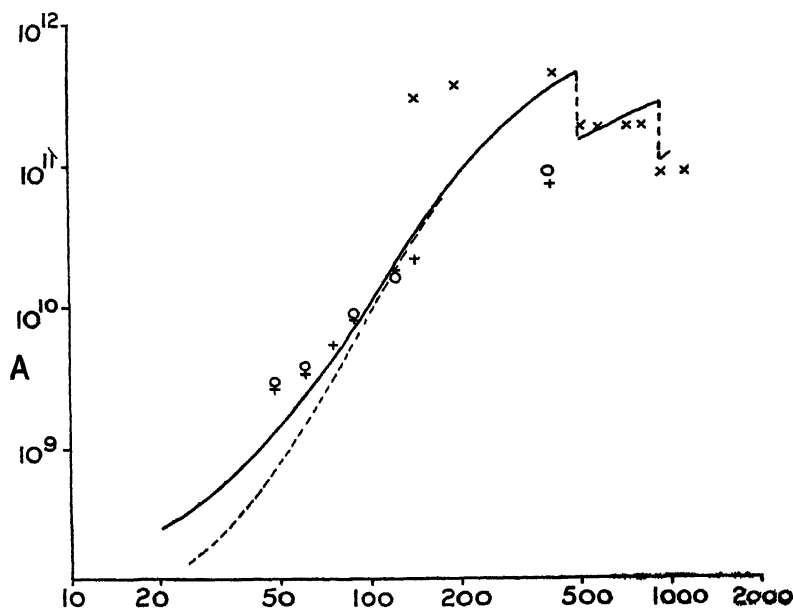


Figure 1

As the silver in the emulsion is the dominant factor in the absorption process which leads to the photographic action, the determination of the effective wave-length of the heterogeneous x-ray beams used was based on absorption measurements in silver in preference to the more usual copper. Comparative measurements showed variations in the measured effective wave-length of about 3 x.u. for the hardest radiations, rising to 10 x.u. for the softest radiation used. The absorption coefficients used were derived from various published data (Compton and Allison, 1935; Jones, 1936; Victoreen, 1943).

Figure 1 shows calculated values for A as defined by equation (6a) for wave-length of 20 to 1000 x.u. plotted on logarithmic scales; the broken line refers to values of A when $(\tau + \sigma)_{\text{AgBr}}$ instead of μ_{AgBr} is used in equation (6a). According to equation (7a) the speed, and similarly, according to equation (11a), the density per roentgen on the linear portion of the density-exposure curve would be proportional to A for radiation of different penetrating powers if η were independent

of the effective wave-length, the other factors being naturally unaffected by the wave-length. The experimental results of exposures to x rays from $\lambda=47$ x.u. to 395 x.u. are included in figure 1, and represent values for the speed for film Y and values for the density per roentgen for film X (see table 2). Berthold's values (1925) for the density divided by arbitrary units of ionization measured with monochromatic beams are included to extend the comparison into the region of the K-absorption edges of silver and bromine. All values were arbitrarily matched to the curve at one point.

Table 2

Film	Marked in figures 1 and 3	Experimental values in terms of
X	+	density per r. (constant d)
Y	⊙	ϵ from density-exposure curves corrected by (8)
—	×	Berthold (1925) density per arbitrary unit of ionization

Figure 1 shows that the wave-length dependence of the photographic action is essentially the same for the different ways of derivation from the density-exposure curves and also the same for both emulsions used, but it also reveals a wide divergence between calculated and measured values. The trend of the measured values compared with the theoretical values for the term A is unmistakable. The values for $\lambda_{\text{eff}} = 47$ x.u. are about twice the theoretical value, whilst

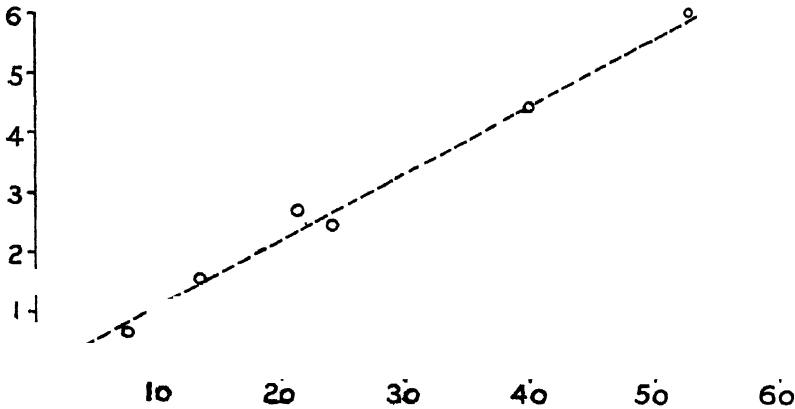


Figure 2

the values for $\lambda_{\text{eff}} = 395$ x.u. are only a quarter, with a definite trend for the intermediate values, differences which are clearly far beyond the possible error estimated above. Berthold's values show the same tendency in the region of short wave-lengths as well as between the absorption edges.

Besides the factor A , the only parameter in equations (7a) and (11a) which may depend on the quality of the radiation used is the quantum efficiency, which was therefore investigated by calculating the number of photons absorbed and

counting the number of developed grains per unit area under the microscope. Equation (11a) was used to calculate the quantum efficiency from observed densities. Both methods give essentially the same results.

Figure 2 shows for film X the number of grains actually counted per absorbed photon against the "effective" frequency as a measure of the average energy of the absorbed x-ray quanta. Figure 2 suggests, as indicated by the dotted line, a linear increase of the quantum-efficiency with the frequency. This means, however, as comparison with equation (2) shows, that the number of grains made developable is proportional to the absorbed energy, independent of the wavelength. The elementary processes leading to these observed quantum efficiencies are not easily assessed, and no theory is attempted at this stage.

Using the Duane-Hunt relation, the observations can conveniently be presented in the form of equation (12), where k is the number of grains affected per kev. of absorbed energy, and, therefore, the reciprocal of k is the energy in kev. necessary to make one grain developable :

$$\eta = 12.35k/\lambda. \quad \dots (12)$$

Inserting (12) in (7) gives

$$\epsilon = Bk\gamma, \quad \dots (13)$$

where

$$B = \frac{\mu_{\text{AgBr}}}{(\tau + \sigma_{\text{a}})_{\text{air}}} 1.14 \times 10^7 \quad \dots (14)$$

and

$$n = N(1 - e^{-Bk\gamma}) \quad \dots (15)$$

For the linear part of the density-exposure curve we may write

$$\frac{d}{\gamma} = 0.4343\alpha_{\text{lg}}Bk. \quad \dots (16)$$

Figure 3 gives calculated values for B , plotted as $\log B$ against $\log \lambda$. It should be noted that the ordinate in figure 3 is drawn to twice the scale used in figure 1, and, therefore, the slope of the experimental values appears to be steeper. The fall in the efficiency for harder radiations found by Berthold does not appear in figure 3, and was obviously due to the values for the absorption coefficients used by Berthold in 1925. The agreement between experimental and calculated values is satisfactory for all series, which are matched to the curve as in figure 1 by using one free parameter, the calculation of which seems desirable and can be attempted by the use of equations (13), (15) and (16). A discussion of the different factors involved may help to assess the accuracy to be expected for the comparison between measured and calculated values in tables 3, 4 and 5.

Determination of the factor k involves the measurement of the quantum efficiency at a certain stage of development, and has to be extrapolated for maximum development, because equation (13) was deduced from a consideration of the absorption in all grains and must, therefore, be referred back to full utilization of all affected grains. Term B includes the absorption coefficients of silver bromide and air. The influence of the divergencies in the various published absorption coefficients has to a certain extent been eliminated by the choice of silver for the absorption measurements in the present investigation ; the absorption

coefficients for air are calculated values for the whole range. The photometric constant defined as the amount of silver per unit area divided by the density can be calculated from the value for the average mass per grain (γ), involving, apart from the measurement of density and mass of silver per unit area, the estimation

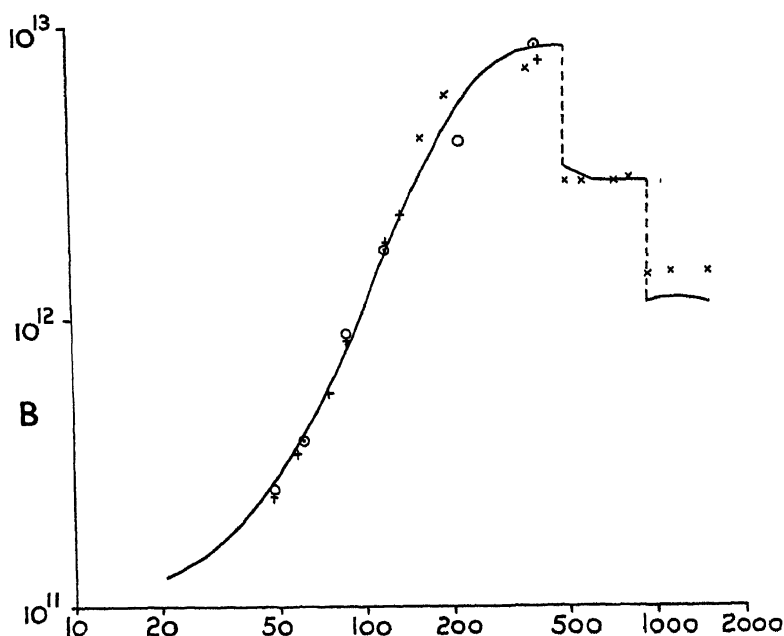


Figure 3

of the average area per grain based on the microscopically measured average diameter per grain, i. e. the square of the diameter and the shape of the grain will influence the result. The composition of the emulsions used from various classes of grain might to a certain extent influence the calculation from basic data, although it does not influence the arbitrary values in figure 3.

Table 3

Film	$\lambda \times v$	ϵ meas	η meas	k grains/ kev	γ calc. gm / grain	Photometric constant calculated (gm / dm ²)	Photometric constant measured (gm / dm ²)
X	395	0.03	2.3	0.073	4.8×10^{-14}	1.0×10^{-3}	1.1×10^{-3}
Y	87	0.73	68	0.48	1.9×10^{-12}	2.8×10^{-3}	3.3×10^{-3}

Table 3 gives, for the two emulsions used, measured values for the factor ϵ and for the quantum efficiency (η), the latter being corrected for maximum development. The values for the number of grains made developable per kev. absorbed are calculated from equation (12) using the corresponding values for the quantum efficiency. The average mass of AgBr per grain (γ) calculated from equation (13)

was used to calculate the photometric constant in grams of silver per sq. dm., no form factor (Arens, Eggert and Heisenberg, 1931) for the influence of the shape of the grains being taken into account. The agreement between calculated and measured values for the photometric constant can be considered satisfactory, and this comparison serves as a test for the validity of the theoretical considerations set out in this paper.

For the calculation of densities for emulsion X from equation (16) in table 4, the relevant factors were taken from tables 1 and 3. In table 5, similar calculations for film Y using equations (15) and (8) are shown, the total number of grains per sq. cm. (*N*) having been calculated from the measured total of AgBr per sq. cm. (table 1) and the value for the average mass per grain (*γ* in table 3). The general agreement for various wave-lengths is shown in figure 3, and only a limited number of values is given in tables 4 and 5.

Table 4
Film X; $0.4343\alpha\bar{a}gk=4.3 \times 10^{-14}$

kv.	Filter	λ effective x.u	<i>B</i>	Exposure <i>r</i>	<i>d</i> meas	<i>d</i> calc
400	2 mm. Sn	60.5	3.8×10^{11}	28.0	0.428	0.458
400	Inherent	74	5.8×10^{11}	27.4	0.683	0.66
190	Thoraeus	119	1.7×10^{12}	4.5	0.335	0.33
190	1 mm Al	208	5.2×10^{12}	0.74	0.184	0.166

Table 5
Film Y; $N=1.37 \times 10^8$

kv	Filter	λ effective x.u	Exposure <i>r</i>	<i>d</i> meas	<i>d</i> calc
190	Thoraeus	119	0.297	1.08	1.08
400	2 mm Sn	60.5	2.04	1.35	1.46
475	4.5 mm. Sn	47	2.75	1.41	1.5

§ 4 DISCUSSION OF RESULTS

In disagreement with the assumption of previous workers (Silberstein and Trivelli, 1930 ; Charlesby, 1940), the theoretical findings and experimental results described above indicate that the speed of the emulsion is proportional to the product $\eta \times \gamma$ (equation 7 *a*), a conclusion of some significance and calling for a brief discussion.

In the well known absorption formula

$$Q = Q_0 e^{-\mu d},$$

the number of photons (Q_0) crossing a grain will be proportional to the area, whilst the fraction absorbed depends on the thickness and absorption coefficient only. Thus proportionality between the number of photons absorbed and the area would exist only in emulsions consisting either of opaque grains or of grains of equal thickness. The highest absorption coefficient for silver bromide met with in this investigation is of the order 1000 (for $\lambda = 1.5 \text{ \AA}$), and taking the diameter of the fairly large undeveloped grain of emulsion Y as $0.7 \times 10^{-4} \text{ cm}$ gives an absorption of about 7% of the photons crossing a grain, thus excluding any assumption of opacity. The undeveloped grain is regarded as a minute silver-halide crystal in modern theories of the latent image (Gurney and Mott, 1938; Mott and Gurney, 1940; Berg, 1943), and there seems to be no reason to assume equal thickness of grains.

L. Silberstein and A. P. H. Trivelli (1930) tested the validity of equation (1c) for a series of grain sizes, using single-layer plates, and claimed good agreement between theory and experiment. Re-examination of their results shows, however, that the spread of the experimental values is too large to allow of a clear decision. A brief discussion of the possible dependence of the quantum efficiency on grain-size is given in a later paragraph.

Integration of equation (4) gives

$$n = N(1 - e^{-\frac{\mu}{\rho} \rho Q}),$$

where the term in the exponent should be equal to the term in the exponent of equation (1c), i.e., we should have

$$\frac{\eta \mu \gamma}{\rho} = a \beta. \quad \dots\dots(17)$$

A. Charlesby (1940) estimated the term a from G. E. Bell's measurements of 1936, and found values ranging from 1.6×10^{-10} to 4.8×10^{-10} for films which seem to be comparable to film Y. From the data given in Bell's paper, μ can be estimated to be 200, η of the order 10, $\gamma = 1.9 \times 10^{-12}$ (see table 3, film Y), and, therefore, $\frac{\eta \mu \gamma}{\rho} = 6 \times 10^{-10}$, in sufficiently good agreement considering the data available as a basis of estimation.

The average energy required to make one grain developable, estimated from the measured quantum efficiencies and corrected for maximum yield from the increase of density for prolonged development (table 3), is 13,000 ev. for film X and 2100 ev for film Y, which, even with due regard to the error in the determination of the quantum efficiency, is higher than the values given for exposures with visible light. F. R. Hirsch (1928) found that emulsions exposed to x rays of 11 \AA . (1100 ev.) show the toe, characteristic of exposures to visible light, at the beginning of the density-exposure curve. Taking this as an indication that the quantum efficiency is below unity for this wave-length and the emulsion used (analysis of Hirsch's published graph indicates a quantum efficiency of approximately one-fourth), the same order for the average energy per grain is obtained. Mott and Gurney (1940) estimate that on the average about 50 quanta of visible light are needed to make one grain developable, an average absorbed

energy of about 200 ev. per grain. S. E. Sheppard (1931), however, comes to the conclusion that on the average from a few hundred up to one thousand quanta are needed for visible light. The efficiency of the photographic action of x-ray exposures might, however, on account of the almost instantaneous release of a great number of electrons, be nearer to the efficiency of visible light of high intensity, where, as the breakdown of the reciprocity law shows, a higher average energy per grain is required. The theoretical aspect is further complicated by the difference in the energies necessary (film X with a small grain needing more than film Y with a larger grain), which, if not due to the preparation of the films, might be explained on Mott and Gurney's theory by a higher rate of recombination due to higher concentration of electrons in a smaller grain. An interesting aspect of the sensitivity of large grains is presented by W. F. Berg (1943) in a discussion of the electrolytic mechanism of development: "A large grain containing a single sensitivity speck will be more sensitive, not only because of its larger target area to light, but also because it contains a larger total number of interstitial ions to begin with".

Measured values for the exposure in roentgens necessary to produce a density of 0.5, plotted against half-value layer in Cu published by G. E. Bell (1936) and E. E. Smith (1943), were compared with values calculated from equation (16) and show reasonable agreement for harder radiations. The minimum at about 0.25 mm. Cu can be explained by the influence of quanta of lower energy than the K-absorption edge of silver which would have been present in the heterogeneous beams used, and tend to raise the amount of x rays needed before the effective wave-length coincides with the absorption edge.

The theoretical considerations presented in this paper should in principle apply to exposures to visible light. A different mathematical form would have to be chosen to account for the high scattering of light in the emulsion, absorption due to dyes and variation in the quantum efficiency for different grains. A rough estimate of the value of the intrinsic sensitivity (β), using equation (17), can, however, be gained in the following way. J. Eggert and W. Noddack (1923) measured the absorption of light in the silver bromide of photographic emulsions and found a total absorption of 10 to 20%. Assuming a total of 4×10^{-4} gm. AgBr per sq. cm. gives a thickness of 0.6×10^{-4} cm., which absorbs, say, 15% of the incident light; an absorption coefficient of $2.5 \times 10^3 \text{ cm}^{-1}$ is deduced. This high value for the absorption coefficient is explained by the fact that the method of calculation automatically takes the multiple scattering in the emulsion into account. Taking the diameter of the undeveloped grain as 10^{-5} cm. and the quantum efficiency (Sheppard, 1931) as 3×10^{-3} gives $\beta = l\eta\mu = 7.5 \times 10^{-5}$. Various workers have arrived at values of 10^{-3} to 10^{-5} for the intrinsic sensitivity.

ACKNOWLEDGMENTS

The author wishes to thank Dr. L. H. Clark and Mr. D. E. A. Jones, B.Sc., for their interest and for their help in the preparation of this paper, and the Medical Officer of Health, Public Health Department, London County Council, for opportunities afforded.

REFERENCES

- ARENS, H, EGGERT, J and HEISENBERG, E, 1931 *Z wiss. Phot.* **28**, 356
 BELL, G. E, 1936 *Brit. J Radiol.* **9**, 578.
 BERG, W. F, 1943. *Trans. Faraday Soc* **39**, 115
 BERTHOLD, R., 1925. *Ann Phys., Lpz*, **4**, 76, 409
 BLAU, M. and ALTENBURGER, K, 1922 *Z Phys* **12**, 315
 CHARLESBY, A, 1940 *Proc Phys Soc.* **52**, 657.
 COMPTON, A. H and ALLISON, S K, 1935. *X-rays in Theory and Experiment.*
 (London: Macmillan)
 EGGERT, J. and NODDACK, W., 1923. *Z Phys* **20**, 299
 GLOCKER, R. and TRAUB, W, 1921. *Phys. Z* **22**, 345
 GURNEY, R W. and MOTT, N F, 1938 *Proc Roy Soc A*, **164**, 151
 HIRSCH, F R, 1928. *J Opt Soc Amer* **28**, 463
 JONES, M T, 1936 *Phys. Rev* **50**, 110
 MOTT, N F. and GURNEY, R W, 1940 *Electronic Processes in Ionic Crystals.*
 (London: Clarendon Press)
 NUTTING, P G., 1913 *Phil. Mag* **26**, 423.
 SHEPPARD, S E, 1931. *Photogr J* **71**, 313
 SILBERSTEIN, L., 1922 *Phil Mag* (6), **44**, 257 and 962
 SILBERSTEIN, L. and TRIVELLI, A P. H, 1930 *Phil Mag* **9**, 787
 SMITH, E E, 1943 *Photogr J* **83**, 363
 VICTOREEN, J A, 1943 *J Appl Phys* **14**, 95

A METHOD FOR OBTAINING SMALL MECHANICAL VIBRATIONS OF KNOWN AMPLITUDE

By D H SMITH,
Woolwich Polytechnic

MS received 4 July 1945

ABSTRACT The theory of a method for measuring small amplitudes of vibration, due originally to Thomas and Warren, is developed, and experiments confirming it are described. The possibility of using the method for producing a standard source of sound is discussed

§ 1. INTRODUCTION

THE amplitude of a vibrating body such as a diaphragm may be measured either by means of a device which depends upon contact with the body or with an optical arrangement involving the use of interference fringes. The best-known amplitude meters of the first class are the Bragg amplitude meter (Bragg, 1919) and the optical lever used by Kennelly (1923).

Neither of these is very accurate. The first, which depends on the cessation of chattering between the diaphragm and a light spring-loaded hammer, is cumbersome, and cannot well be mounted behind the diaphragm without interfering with the driving arrangements. Kennelly's optical lever is fragile, and needs a powerful arc as a source of light, while its accuracy is low for amplitudes

less than about 10 microns. In particular, its usefulness diminishes towards higher frequencies, where a large particle velocity, and, therefore, a high intensity of sound, are obtained with a small amplitude.

Interference fringes were used by Webster (1919), who photographed the fringes produced when a mirror attached to the vibrating body was made one of the mirrors of a Michelson interferometer. A much simpler method was devised by Thomas and Warren (1928), who have described the appearances presented by thin film interference fringes of the Newtonian kind, formed between a reflecting surface at rest, and one set in vibration with a small amplitude. Their description is, however, not quite accurate in several particulars, and the qualitative explanation of the phenomena which they gave appears to have led them to erroneous quantitative conclusions.

In the present paper, the theory of the fringes observed by Thomas and Warren is examined in detail, and a source of sound is described whose amplitude can be accurately measured by their use.

§ 2 THE FORMATION OF INTERFERENCE FRINGES BY REFLECTION FROM THE SURFACES OF A THIN FILM, ONE OF WHICH IS VIBRATING

Let AB and CD be two close reflecting surfaces of any form, not in contact at any point, and consider the interference of the monochromatic beams OS and OPQS reflected from these surfaces. Let the vibrations in the reflected

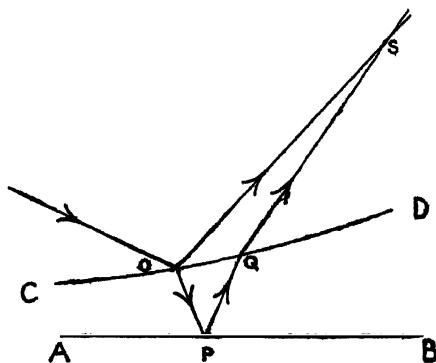


Figure 1.

beams be $y_1 = a_1 \cos \omega t$ and $y_2 = a_2 \cos (\omega t - \delta)$, where a_1 and a_2 are the amplitudes and δ is the phase difference. The resultant is, in complex form,

$$y_1 + y_2 = \exp. j\omega t [a_1 + a_2 \exp. -j\delta].$$

The intensity, obtained by multiplying by the conjugate complex quantity, is

$$\begin{aligned} I &= (a_1 + a_2 \exp. -j\delta)(a_1 + a_2 \exp. j\delta) \\ &= a_1^2 + a_2^2 + a_1 a_2 (\exp. -j\delta + \exp. j\delta) \\ &= a_1^2 + a_2^2 + 2a_1 a_2 \cos \delta \end{aligned} \quad \dots\dots (1)$$

Let e_0 be the distance between the surfaces at the point O, when they are at rest. Then with sensibly normal incidence

$$\delta = \frac{4\pi e_0}{\lambda},$$

where λ is the wave-length. Let the lower plate vibrate in a sense normal to the path of the light, its motion being $\xi = \xi_0 \cos pt$. The separation of the surfaces at O becomes

$$e_0 + \xi \quad \text{and} \quad \delta = \frac{4\pi}{\lambda} (e_0 + \xi_0 \cos pt).$$

Let

$$\frac{4\pi\xi_0}{\lambda} = \epsilon.$$

Then the intensity at O is

$$\bar{I} = a_1^2 + a_2^2 + 2a_1a_2 \cos(\delta + \epsilon \cos pt).$$

The mean intensity at this point over a period is

$$\begin{aligned} \bar{I}_0 &= \frac{1}{T} \int_0^T I_0 dt \\ &= \frac{1}{T} \int_0^T [(a_1^2 + a_2^2) + 2a_1a_2 \cos(\delta + \epsilon \cos pt)] dt \\ &= (a_1^2 + a_2^2) + \frac{2a_1a_2}{T} \int_0^T \cos(\delta + \epsilon \cos pt) dt \quad \dots\dots (2) \end{aligned}$$

The integral in (2) is

$$I_1 = \int_0^T \cos(\delta + \epsilon \cos pt) dt = \cos \delta \int_0^T \cos(\epsilon \cos pt) dt - \sin \delta \int_0^T \sin(\epsilon \cos pt) dt \quad \dots\dots (3)$$

The second integral in equation (3) is zero, since by writing $\tau = t - T/2$ we see that $\int_{-T/2}^{T/2} = - \int_0^{T/2}$. If we write $pt = \phi$, the first term becomes

$$\begin{aligned} \cos \delta \int_0^T \cos(\epsilon \cos pt) dt &= \frac{\cos \delta}{p} \int_0^{2\pi} \cos(\epsilon \cos \phi) d\phi \\ &= \frac{\cos \delta}{p} \int_0^{2\pi} \left\{ 1 - \frac{\epsilon^2 \cos^2 \phi}{2!} + \frac{\epsilon^4 \cos^4 \phi}{4!} - \frac{\epsilon^6 \cos^6 \phi}{6!} + \dots \right\} d\phi. \end{aligned}$$

Using the well-known reduction formula

$$\int_0^{2\pi} \cos^{2n} \phi d\phi = \frac{1 \cdot 3 \cdot 5 \dots (2n-1)}{2 \cdot 4 \cdot 6 \dots 2n} \cdot 2\pi,$$

this becomes

$$\begin{aligned} &\frac{2\pi \cos \delta}{p} \left\{ 1 - \frac{\epsilon^2}{2^2} + \frac{\epsilon^4}{(2 \cdot 4)^2} - \frac{\epsilon^6}{(2 \cdot 4 \cdot 6)^2} + \dots \right\} \\ &= \frac{2\pi \cos \delta}{p} \cdot J_0(\epsilon) = T \cdot J_0(\epsilon) \cos \delta, \end{aligned}$$

where J_0 is the Bessel function of zero order. Substituting this result in equation (2), we have finally

$$I_0 = (a_1^2 + a_2^2) + 2a_1a_2 \cdot J_0(\epsilon) \cos \delta \quad \dots\dots (4)$$

$$(a_1^2 + a_2^2) + 2a_1a_2 J_0\left(\frac{4\pi\xi_0}{\lambda}\right) \cdot \cos\left(\frac{4\pi e_0}{\lambda}\right). \quad \dots\dots (5)$$

With a given value of ξ_0 , which is the same at all points, the intensity varies across the field of view with e_0 and a system of thin-film interference fringes will be formed, whose configuration will depend on the shape of the reflecting

surfaces. In the particular case in which we shall be interested, that of fringes formed by reflection from the surfaces of a lens and a flat plate, the usual system of Newton's rings will appear. The contrast between the bright and dark fringes—that is, the visibility—will depend on the value of $J_0\left(\frac{4\pi\xi_0}{\lambda}\right)$. Thus, when $\xi_0=0$, so that $J_0\left(\frac{4\pi\xi_0}{\lambda}\right)=1$, the fringes will vary in intensity between $(a_1-a_2)^2$ in the dark fringes and $(a_1+a_2)^2$ in the bright fringes. But for those values of ξ_0 which make $J_0\left(\frac{4\pi\xi_0}{\lambda}\right)=0$, the intensity will be uniform and equal to $a_1^2+a_2^2$ over the whole field, and for these critical values of ξ_0 the fringes will disappear.

The first few values of ϵ/π and, therefore, of $4\xi_0/\lambda$, for which $J_0(\epsilon)=0$, are shown in the first column of table 1, while the second column shows the corresponding critical amplitude as a fraction of the wave-length of the light used.

Table 1
Critical values of the amplitude

$\frac{\epsilon}{\pi} = \frac{4\xi_0}{\lambda}$	$\frac{\xi_0}{\lambda}$	$\frac{\epsilon}{\pi} = \frac{4\xi_0}{\lambda}$	$\frac{\xi_0}{\lambda}$
0.7655	0.1914	5.7522	1.4380
1.7571	0.4393	6.7519	1.6880
2.7546	0.6886	7.7516	1.9379
3.7534	0.9383	8.7514	2.1879
4.7527	1.1882	9.7513	2.4378

In passing through the value zero, $J_0(\epsilon)$ changes sign, and so, therefore, does the second term in equation (4), so that as the amplitude is gradually increased through a critical value, the fringes disappear and then reappear with the dark and bright bands interchanged in position.

These conclusions are entirely borne out by experiment. Thus, if a well-defined set of fringes is formed with the lower plate at rest, and it is then set in vibration with a very small amplitude, much less than the first critical amplitude, the fringes diminish slightly in visibility. If the amplitude is gradually increased, the contrast diminishes still further, and when the first critical amplitude is reached, the fringes disappear and the field becomes uniformly illuminated. A further increase of amplitude causes the fringe system to reappear with bright and dark bands interchanged. This is repeated as each critical amplitude is reached. The change in the position of the fringes in passing through a critical amplitude is a valuable feature of the effect, as it enables the amplitude to be set for uniform illumination with a precision approaching that obtainable with a good photometer.

The figures in table 1 show that $J_0(\epsilon)$ vanishes for values of ϵ/π , which differ by very nearly unity, and the difference approaches unity closely as ϵ/π increases.

Therefore the successive disappearances of the fringes occur with amplitudes which differ by nearly $\lambda/4$, but the critical amplitudes themselves are not exactly odd multiples of $\lambda/8$, as stated by Thomas and Warren. The theory shows that with high amplitudes the fringes become less distinct, and it becomes slightly more difficult to adjust the amplitude to a critical value. In practice, a_1 and a_2 are very nearly equal, and if they are exactly equal, equation (4) becomes

$$\bar{I}_0^2 = 2a^2[1 + J_0(\epsilon) \cos \delta].$$

For a given value of ξ_0 , the visibility of the fringes is then

$$V = \frac{\bar{I}_{\max} - \bar{I}_{\min}}{\bar{I}_{\max} + \bar{I}_{\min}} = J_0(\epsilon).$$

Negative values of V correspond, of course, to the interchange of the bright and dark fringes on passing through a critical amplitude. The visibility, starting at unity when ξ_0 is zero, reaches maximum values of about 0.4, 0.3, 0.25 and 0.21 between the first and second, second and third, third and fourth, fourth and fifth critical amplitudes, and thereafter decreases much more slowly, so that between the twentieth and twenty-first disappearances it is still 0.1, and between the thirtieth and the thirty-first 0.078. Thus the fringes are still clearly visible even in the neighbourhood of the thirtieth disappearance, and, as described later, the amplitude can still be set accurately to the critical value.

§ 3. EXPERIMENTAL

The conclusions reached in § 2 have been verified with the apparatus shown (figure 2). The vibrator is a flat aluminium-alloy diaphragm (A), 0.025 in.

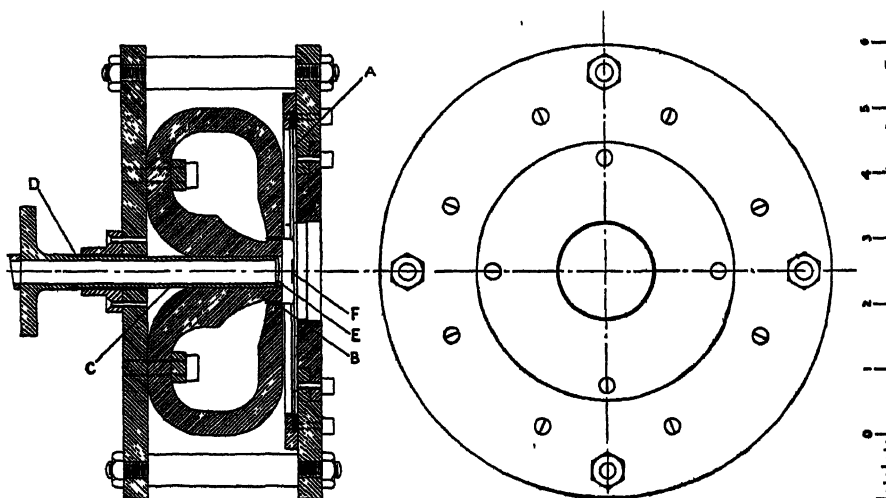


Figure 2

thick and 4.75 in. in diameter, which is tightly clamped to the heavy brass front plate of the instrument by means of a steel ring. Several layers of cartridge paper smeared with hard vacuum grease are interposed between the diaphragm

and the ring, while a single layer of paper is placed between the diaphragm and the front plate; this is done because in one application of the instrument it is desired to have a gas-tight joint here. The free portion of the diaphragm is 4.5 in. in diameter, and it is separated from the front plate by a distance of 0.005 in. The diaphragm, after being cut from the sheet, was flattened by clamping it between flat steel plates in a large screw press and heating it to about 300° c. The diaphragm is maintained in vibration by means of a "moving coil" drive. The coil (B) consists of two layers, each of 18 turns, of 38 S.W.G. silk-covered copper wire wound on a stiff paper former one inch in diameter, which is cemented to the diaphragm. It moves in the annular gap of a large claw-shaped permanent magnet, of the kind formerly used in moving-coil loudspeakers. The depth of the gap is 0.25 in., while the axial length of the coil is 0.18 in., and the coil is wound in such a position on the former that it is symmetrically situated in the gap. It may be assumed therefore that, when vibrating with small amplitude, the coil remains in a uniform magnetic field, so that a sinusoidal current will produce strictly sinusoidal vibrations. The magnet itself is clamped to a heavy brass back-plate, and the front plate bearing the diaphragm is carried on four stout pillars rising from the back plate. A hole 0.5 in. in diameter is bored through the central soft-iron pole-piece of the magnet, and a close-fitting brass tube (C), whose internal diameter is 0.375 in., can be advanced through it by means of a screw (D) cut on the outside of the tube and working in a nut which is bolted to the back plate. At the end of the tube is mounted a plano-convex lens (E), the radius of curvature of whose front surface is 200 cm. The lens is mounted on a brass plate which can be tilted by means of three small spring-loaded screws; for the sake of clarity, this arrangement is not shown in the diagram. At the centre of the diaphragm, opposite the lens, is cemented a circular microscope cover-slip (F), whose back face is lightly ground and coated with optical black to abolish unwanted reflections. With the lens close to the glass plate, but not quite touching it, Newton's rings are formed by reflection at their surfaces when monochromatic light is projected down the tube. The screw has fifty threads per inch, so that the distance between the lens and the plate can be delicately adjusted. The tilting screws are manipulated with a long screw-driver passed up the tube, and with a little practice it is easy to bring the centre of the rings to the middle of the field of view, while by turning the tube, the separation of the lens and the plate is adjusted to be about 10 microns. The fringes are viewed through the tube with a telemicroscope.

An aluminium cover, not shown in the diagram, encloses the rear part of the apparatus. As the whole instrument is massively constructed and the moving parts are accurately fitted, it is scarcely affected by shock, and the only effect of tapping the case is to produce a slight quiver of the fringes, which is due to free vibration of the diaphragm.

Observations were made at a number of frequencies of the R.M.S. currents required to produce disappearance of the fringes, i.e. at the critical amplitudes. The current was measured with a Unipivot galvanometer with a square-law scale and a series of vacuum thermojunctions whose ranges overlapped, so that the current readings at each critical amplitude could be made with nearly the same

accuracy. The thermojunctions were calibrated with direct current, using a direct-reading potentiometer and standard resistances.

The observations were made with sodium light, whose lack of monochromaticity introduces no appreciable error into the results.

§ 4 RESULTS AND DISCUSSION

It was found that the amplitude could be set to a critical value with an accuracy of about one-half per cent, as judged by the agreement between the values of

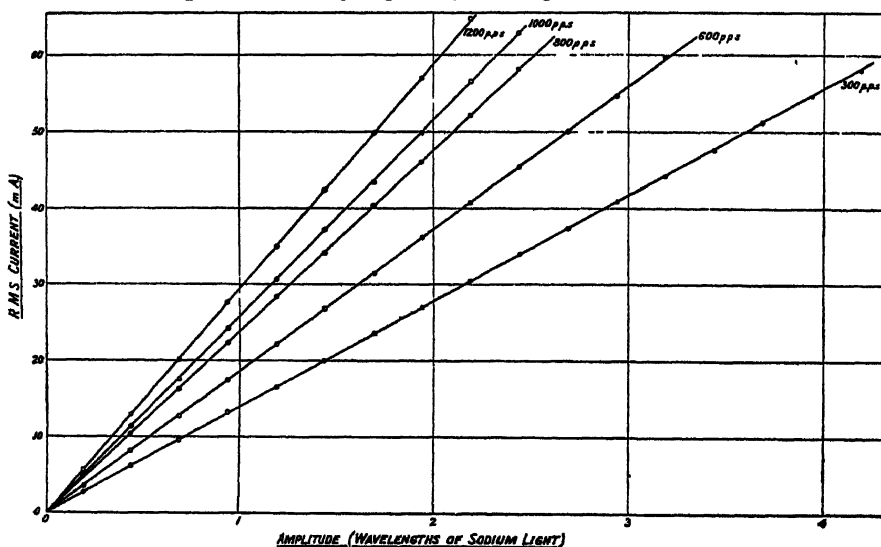


Figure 3.

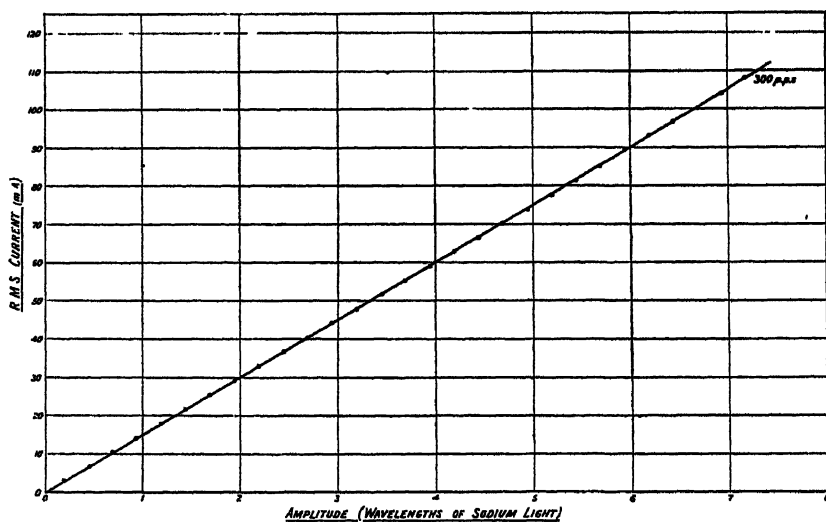


Figure 4

the current with successive settings. At the higher amplitudes, the falling-off in visibility is more than offset by the fact that a given change in amplitude is brought about by a smaller fractional change of current than at lower amplitudes.

The results at five frequencies are shown graphically in figure 3, where the current for which the fringes disappear is plotted against the corresponding critical amplitude given by the calculation of § 2. Each point represents the mean of four observations. In each case the curve relating current and amplitude is accurately a straight line passing through the origin, thus confirming the theory.*

It is evident, from the accuracy with which the points lie on a straight line, that when the best straight line through them has been found, the amplitude can be set to any assigned value within the range by adjusting the current, with an error not greater than 1 in 500. In the case of the observations at 300 c./sec., the equation of the straight line through the origin and the centroid of the points

$$\mathcal{I} = 13.900(\xi/\lambda),$$

where \mathcal{I} is the current, and the maximum deviation (on the current axis, since the amplitude values are pre-assigned) which occurs at the seventeenth disappearance, with a current of 58 ma., is 0.2 ma.

In figure 4 is shown the result of observations extending to the thirtieth disappearance ($\xi_0 = 4.38 \mu$). It would appear justifiable to extrapolate over double this range. The amplitude would then be about 8μ , and there is no reason to suppose that the law connecting current and amplitude would cease to be a linear one.

The instrument shown in figure 2 has been designed primarily for use in a new tube method for measuring the velocity of sound, but with slight modification it constitutes a standard source of sound, whose amplitude can be set to have a known absolute value with high accuracy. It has been described here in the form with which the observations for figures 3 and 4 were made, but a more compact pattern has been built, in which an axially magnetized cylindrical magnet is used. In order to make it suitable for use as a standard source, it is fitted with a thinner front plate, and the diaphragm carries a composite piston of balsa wood and aluminium, whose front face is coplanar with the front plate. When it is mounted in an extensive baffle in an acoustic chamber, it produces a sound field whose distribution can be calculated. The R.M.S. displacement velocity at a point on the axis l cm. in front of such a piston, of radius r , is given by

$$\dot{\xi} = \frac{\dot{\xi}_0}{\sqrt{2}} \left\{ \left(\frac{l}{\alpha} \right)^2 + 1 - \frac{2l}{\alpha} \cos \frac{2\pi}{\lambda_1} (\alpha - l) \right\}^{\frac{1}{2}},$$

where $\dot{\xi}_0$ is the peak displacement velocity of the piston, λ_1 is the wave-length of the sound, and $\alpha = \sqrt{l^2 + r^2}$ (West, 1932). Table 2 shows the displacement velocity on the axis of the piston, of radius 2 cm., at a distance of 1 metre, together with the approximate intensity in db. above 10^{-16} w. per cm², when the amplitude of the piston is 4μ .† This corresponds to about the thirtieth critical amplitude

* The apparent discrepancy in the slope of the line per 1000 c/sec. is due to the fact that some adjustments were made to the apparatus between this set and the other sets of readings.

† A sound has an equivalent loudness of n B.S. phons when it has the same loudness (judged under specified conditions) as a standard tone in the form of a plane sinusoidal wave at 1000 c/sec., whose intensity level is n db. above a reference level corresponding to an R.M.S. sound pressure of 0.0002 dynes per cm². This corresponds very closely, in air, to an average power of 10^{-16} w. per cm², and the intensities above have been referred to this level, to obtain a rough comparison with speech.

with mercury green light ($\lambda = 5461 \times 10^{-8}$ cm.) The R.M.S. displacement velocity corresponding to the reference level is about 5×10^{-6} cm./sec at 1000 c./sec. As ordinary speech corresponds to an intensity level of about 40 db. at 1 metre, the loudness is adequate for acoustic measurements without extrapolating from calibration curves such as figure 4, except at the lowest frequencies.

Table 2
Intensity of sound from the diaphragm at a distance of 1 metre

f (c/sec)	$\frac{1}{\lambda} \times 10^4$ (cm/sec)	Intensity (db above 10^{-16} w per cm ²)
100	0.654	22.5
500	16.37	50.3
1000	65.4	62.5
2000	262.0	74.4

Experiments are in progress on the use of the source for calibrating microphones, and for other acoustic measurements.

Drysdale (1939) emphasized the need for a standard source of sound, and suggested means by which a spherical source might be constructed, but he pointed out that the chief difficulty would lie in the measurement of the amplitude, which would need to be of the order of 0.2μ . This is well within the scope of the present method.

ACKNOWLEDGMENTS

I am indebted to the Governing Body and the Principal of the Woolwich Polytechnic for providing facilities which enabled this work to be done. I have also to thank High Duty Alloys, Ltd., for presenting a quantity of aluminium-alloy sheet for diaphragms when war-time restrictions made it difficult to obtain this material, and to Darwins, Ltd., for providing some magnets in similar circumstances.

REFERENCES

- BRAGG, W. H., 1919. *Engineering*, 107, 777.
 DRYSDALE, C. V., 1939. *Proc. Phys. Soc.* 51, 359.
 KENNELLY, A. E., 1923. *Electrical Vibration Instruments* (New York, ed by D. C. Jackson and E. R. Reddick).
 THOMAS, H. A. and WARREN, G. W., 1928. *Phil Mag* 5, 1125.
 WEBSTER, A. G., 1919. *Proc. Nat. Acad. Sci., Wash.* 5, 179.
 WEST, W., 1932. *Acoustical Engineering*, 69.

THEORETICAL INVESTIGATION ON TELE- PHOTO LENSES

By T. SMITH, F.R.S.,
Teddington

Paper read to the Optical Group 30 April 1945

ABSTRACT. The theory of thin lenses is employed to investigate possible forms of telephoto lenses consisting of two widely separated thin components. From a consideration of the shapes of the unit surfaces it is shown that permissible forms have a limited range on either side of those in which both components are aplanatically corrected; on the one side the lenses tend to present their more convex aspects to one another, on the other side their more concave aspects. On taking other conditions into account it is found that heavy figuring is required except in two cases. In one of these the components are nearly aplanatically corrected, and in the other they are strongly concave towards one another. The former construction has the advantage in offering less strongly curved surfaces. It is shown that a design of this class, consisting of two ordinary cemented doublets each made from the same kinds of glass, has stable characteristics.

The formulae used in the investigation are given in an appendix

§1 INTRODUCTION

DURING the period of the war the focal lengths of photographic lenses used for reconnaissance have been progressively increased, and to meet this requirement large-scale copies of telephoto lenses designed to meet somewhat different conditions have been made and used extensively. In view of the importance of securing the best possible photographs it nevertheless appeared desirable to carry out a theoretical investigation on telephoto lenses to see whether improvements were likely to be attainable, bearing in mind the work for which these lenses are required. Briefly, the requirements are the attainment of the finest possible definition over a field of angular diameter about 25° , and the comparative unimportance of distortion. For some other uses to which telephoto lens are put, the elimination or reduction of distortion is important, and it may be that to attain this, something which is important in a lens for aerial photography is sacrificed in the normal type of telephoto lens.

It is of great importance that the photographs obtained with these lenses should suffer as little as possible from loss of contrast. With designs otherwise equally good, this is most likely to be secured by a lens with the smallest possible number of glass-air surfaces. It is therefore appropriate to develop the theory in the expectation that neighbouring lens surfaces will be cemented together if satisfactory definition with a suitable aperture can be obtained with this construction. The initial assumption is therefore made that the telephoto combination will consist of two lenses, the leading one of positive power and the other of negative power, each preferably composed of cemented components. The work proceeds on the basis that each of these lenses may, in the first place, be treated as of negligible thickness. As the combination must be chromatically corrected for both position of focal plane and size of image, it is essential that each cemented lens should be

corrected independently for colour. As the lenses are thin this involves only one condition for each lens.

Correction of the combination is secured by imposing suitable conditions on each of the two components. The problem in fact is reduced to the design of two thin lenses. To follow the theory it is therefore important to bear in mind certain properties of thin lenses. For convenience the power of the thin lens is taken as unity.

§ 2 PROPERTIES OF THIN LENSES

Since this lens is to be achromatic at least two glasses with different ν values must be used. The glass with the greater ν is used for an element of positive power, that with the smaller ν for an element of negative power. The algebraic sum of these powers is unity. Let N be the sum of their absolute values. Then the lens will be achromatic if

$$N = \frac{\nu_1 + \nu_2}{\nu_1 - \nu_2}.$$

So far as this condition is concerned it is immaterial whether each kind of glass is represented by a single element or by several elements, whether components are cemented or successive surfaces have different curvatures where they touch on the axis; the order in which the elements are arranged and their shapes are also of no significance. N can therefore be regarded as a current variable when monochromatic aberrations are considered; it will only be necessary at some stage to refer to a catalogue of optical glasses to verify that types giving the value of N desired are available. Negative as well as positive values of N may be admitted, the negative sign indicating that the glass which it was expected would have the smaller value of ν must, in fact, have the greater value. No loss of generality is now involved in supposing that with the definition of N already given ν_1 is the constringence of the glass of lower refractive index, conventionally referred to in what follows as crown glass, and ν_2 that of the glass of higher refractive index, to be referred to as the flint glass. The particular meaning assigned to crown and flint must be kept in mind. For instance, the "crown" may prove to be a glass described in the glass catalogue as a light flint, and the "flint" as a dense barium crown. In such a case the value of N would be negative.

The first-order monochromatic aberrations are usually given as

Spherical Aberration,
Coma,
Astigmatism and Curvature,
Distortion.

Curvature belongs to a different group of aberrations from the other four, and is measured by the Petzval sum for the lens. This is an invariant, independent of the order of the components and their shapes. In a thin lens of two glasses it is a linear function of N . For a thin lens the aperture stop is presumed to be placed in contact with the lens itself. Taking the aberrations in the reverse order to that given above (for thin lenses at least this reversed order is the simpler of the two), the distortion is identically zero, and the astigmatism is invariable and has the same value for all lenses irrespective of their construction. The astigmatism and curvature together determine the shapes of the focal surfaces for

sagittal and tangential lines, they are represented in section in the centre of figure 1 by the circular and oval curves. In any satisfactory instrument for photographic use the lenses must be combined in such a way that the focal surfaces, instead of having these most unsuitable shapes, are substantially coincident over the angular field to be covered.

The two remaining aberrations are under control in different ways. The essential requirement for the elimination of coma is to secure the correct shape for the unit surfaces of the lens, and for the removal of spherical aberration to ensure equality in focal length for rays through different parts of the lens aperture. Some of the means at our disposal for satisfying these conditions in cemented lens combinations are illustrated in figures 1 and 2.

The lenses shown in figure 1 are externally symmetrical—all the external surfaces have the same curvature. Starting from the doublet at the top of the

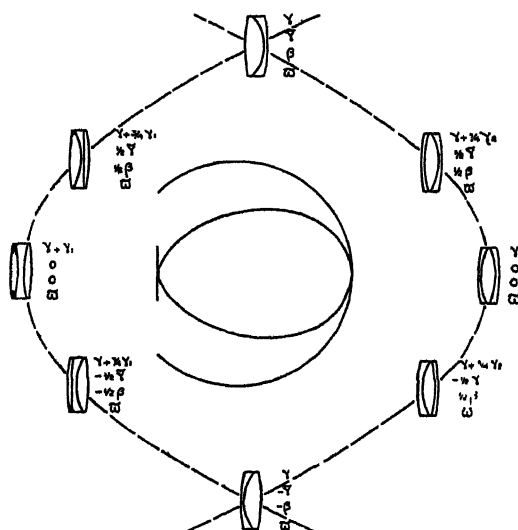


Figure 1. Aberrational coefficients of representative lenses having the same focal surfaces

diagram, with the crown lens leading, the parabolic track on the left leads through forms in which the crown lens is divided, the value of the power of the crown glass transferred from the left to the right side of the flint lens being proportional to the distance descended in the diagram. At the extreme left is shown the symmetrical form in which the two crown components are of equal power. Thereafter the power of the crown component on the left diminishes until at the bottom of the diagram we reach the doublet form with the flint component leading. Lenses could be constructed on the parabola beyond the doublets, but they are less desirable forms, as the curvatures of the internal surfaces are greater than in other forms. The lenses on the other parabola indicate a second series in which the flint lens is divided instead of the crown. Other forms can be constructed in which both flint and crown are divided, so that there are at least four components, but these will not be used in manufacture if simpler forms prove adequate. The aberrational properties of all these externally symmetrical lenses

and of others which can be derived from them involve six coefficients which are functions only of N and of the refractive indices of the glasses. Of these coefficients, ω , the Petzval sum, enters with equal weight in all forms. The next coefficient, β , has weight proportional to the internal asymmetry of the lens, and has its sign reversed if the lens is reversed. As β denotes the sum of the curvatures of the unit surfaces of the lens, this change of sign on reversal is to be expected, as is also the zero coefficient of β for the completely symmetrical forms. The coefficient $\tilde{\gamma}$ behaves in a similar way to β for reasons which will become clear from consideration of a relation to be given later. The ratio of $\tilde{\gamma}$ to β is, in fact, invariable for glasses of given refractive indices. $\tilde{\gamma}$ and β are quadratic in N and vanish when there is only one kind of glass, i.e. when $N = \pm 1$. γ , γ_1 and γ_2 are cubics in N . The roots of $\gamma_1 = 0$ are 1, -1, -1, and those of $\gamma_2 = 0$ are 1, 1, -1; $\gamma = 0$ has only one real root. The weight of γ does not change, but, on proceeding in either direction from the symmetrical forms, $\gamma + \gamma_1$ and $\gamma + \gamma_2$ are diminished by $\theta^2 \gamma_1$ and $\theta^2 \gamma_2$ as the case may be, where θ is the weight given to β and $\tilde{\gamma}$. All the gamma coefficients concern the variation of power with aperture. Of all six coefficients γ is the only one which is modified if the refracting surfaces are not spherical, so that figuring a lens corresponds to an alteration in γ . The method followed in this investigation is to determine how much figuring of each lens is required when all the remaining conditions are satisfied. If no figuring is needed the surfaces are all spherical.

The significance of ω has been given for the focal surfaces, but it is equally important for the unit surfaces. The unit surfaces are free from astigmatism, but the curvature of the first exceeds that of the second by ω . This result again is independent of the order, the division and the shapes of the component lenses.

The aberration known as coma is essentially a variation of magnification with the zone of the lens through which the light passes. To maintain this magnification constant the ratio of the distance along any ray between the second unit surface and the image to that between the first unit surface and the object must be constant. As the difference in curvature of the unit surfaces is fixed, it is obvious that coma can be varied by changing the curvature of both unit surfaces together. If, instead of being zero, as in figure 1, the sum of the curvatures of the external surfaces of the lens is made equal to ρ by increasing the curvature of every refracting surface by $\frac{1}{2}\rho$, the increment in the sum of the curvatures of the unit surfaces is $\rho(1 + \omega)$, so that if this sum were initially $\theta\beta$ it would become $\theta\beta + \rho(1 + \omega)$ after bending. Figure 2 shows a number of forms of lens, both doublets and triplets, bent by a number of equal increments, with the corresponding shapes of the unit surfaces in the last line. In the central column the triplets are both symmetrical, and the curvatures of the unit surfaces equal and opposite. The positions that would be occupied by cemented doublets with equal and opposite curvatures for their external surfaces are indicated by the asterisks in the first and fourth rows.

From the explanation just given it can be readily shown that the condition for freedom from coma when the magnification is m is that the sum of the curvatures of the unit surfaces is $M(2 + \omega)$, where

$$M = \frac{1 + m}{1 - m},$$

i.e. it is

$$\theta\beta + \rho(1 + \varpi) = M(2 + \varpi).$$

When the coma has been removed in this way, the condition that the lens should be free from spherical aberration is

$$\gamma + \gamma_1 - \theta^2\gamma_1 + 2\rho\theta\tilde{\gamma} + \rho^2(1 + 2\varpi) = M^2(5 + 2\varpi),*$$

the lens being a triplet with the crown component divided in the ratio $1 + \theta : 1 - \theta$. If the flint is divided instead of the crown, γ_2 must be written in place of γ_1 . At the moment it is only necessary to note that there are two degrees of freedom

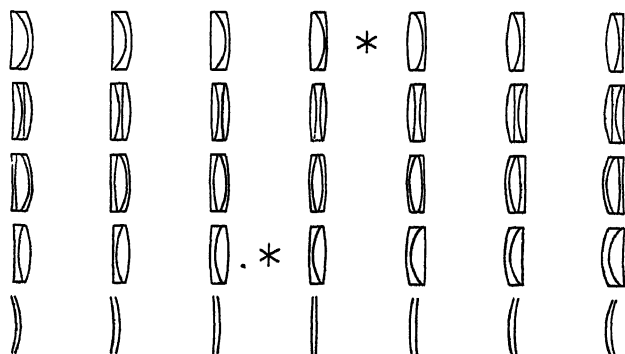


Figure 2 Effect of bending lens on curvatures of unit surfaces

and that in general two conditions can be satisfied. As one of the equations is quadratic it will sometimes happen that the solutions are imaginary.

Returning now to the list of aberrations, and remembering that curvature of field has to be dealt with separately, there are four conditions to be satisfied by the four degrees of freedom in the two thin lenses. But for the reason already given the condition for freedom from distortion will be disregarded and one degree of freedom retained to be used in whatever manner seems most advantageous. The free variable secured in this way will be denoted by ϕ .

§3 THE FUNDAMENTAL STRUCTURE OF THE TELEPHOTO LENS

Before proceeding further it is necessary to consider what the powers of the two component lens groups are to be, and how far they are to be apart. The Petzval sum for the complete lens, which is to vanish, is the sum of those for the components. The ratio of the Petzval sum to the power of a thin lens is only variable to a small extent, so that the powers of the two components will tend to be in a fixed ratio, which is approximately -1 . Suppose that the powers of the two are a and $-ma$. If the power of the combination is unity, the separation t satisfies

$$a - ma + tma^2 = 1$$

and the overall length from the first lens to the focal plane is

$$1 + t - ta$$

* As the value of the power does not involve any distinction between object and image spaces, this condition should not change when the lens is reversed, i.e. when ρ and M are changed in sign. It follows that $\tilde{\gamma}$, like β , changes in sign, but not in magnitude, when the lens is reversed.

or

$$1 - \frac{m}{4} + \frac{1}{m} \left(\frac{1}{a} - 1 + \frac{m}{2} \right)^2.$$

It follows that, for minimum overall length,

$$a = \frac{2}{2-m}, \quad t = \frac{2-m}{4}.$$

In all telephoto lenses of minimum length the separation is therefore half the focal length of the positive component, and the back focal length is half the focal length of the complete lens. It has been customary to describe a telephoto combination with the focal length double the back focal length as a two-times system. In this sense all telephoto lenses of greatest compactness are two-times systems, and this limits the value of the description. An alternative meaning for a two-times system is one in which the focal length of the combination is twice that of the positive component. This will be the case with $m=1$. The systems to be considered here are two-times telephoto lenses in both senses.

§4 UNIT SURFACES OF THE COMPONENTS

With the two components of numerically equal powers, it is appropriate in the first place to assume that they will be made from the same glasses and, therefore, have numerically equal Petzval sums. For a preliminary exploration it may be assumed that the value of the Petzval sum for a thin lens is 0.75 times the power. One form of construction which satisfies the conditions laid down consists of components aplanatically corrected for the magnification at which they are operating. It is clear that this secures freedom from coma and spherical aberration, and under these conditions the astigmatic contributions, which are proportional to the powers of the thin lenses, are also additive, so that both first-order astigmatism and curvature are eliminated. Since for the positive component $m=0$, and for the negative $m=2$, the corresponding values of M are 1 and -3 , and the sums of the curvatures of the unit surfaces of these two components are required to be

$$2 \times 1 \times (2 + .75) = 5.5$$

and

$$(-2) \times (-3) \times (2 + .75) = 16.5.$$

Their differences are $2 \times .75$ and $-2 \times .75$ respectively, or 1.5 and -1.5 , giving for the curvatures of the four unit surfaces in order the values

$$3.5, 2.0, 7.5, 9.0.$$

When the components are not aplanatic, it may be shown that the sum of the curvatures of the first should be increased by $\frac{1}{2}\phi$ and that of the second by -2ϕ . The curvature of the second unit surface of the complete lens must be unity, since the focal length is to be the same for all rays incident parallel to the axis. Figure 3 shows the dependence of all these curvatures on ϕ .

It is profitable to exhibit the variation of the unit surfaces with ϕ in another way, shown in figure 4. Consider first the case marked $\phi=0$. On the extreme left two rays incident parallel to the axis are represented by full lines. They travel in this direction until they meet the first unit surface of the positive lens. (Owing to the sign of the difference in the curvatures of the unit surfaces, the first

unit surface appears on the right of the second, when the light travels from left to right, if, as here, the surfaces are in contact on the axis.) The rays then converge from the points of the second unit surface at the same distance from the axis

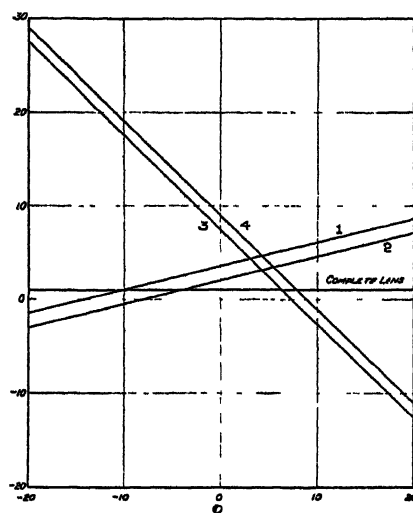


Figure 3 Curvatures of unit surfaces

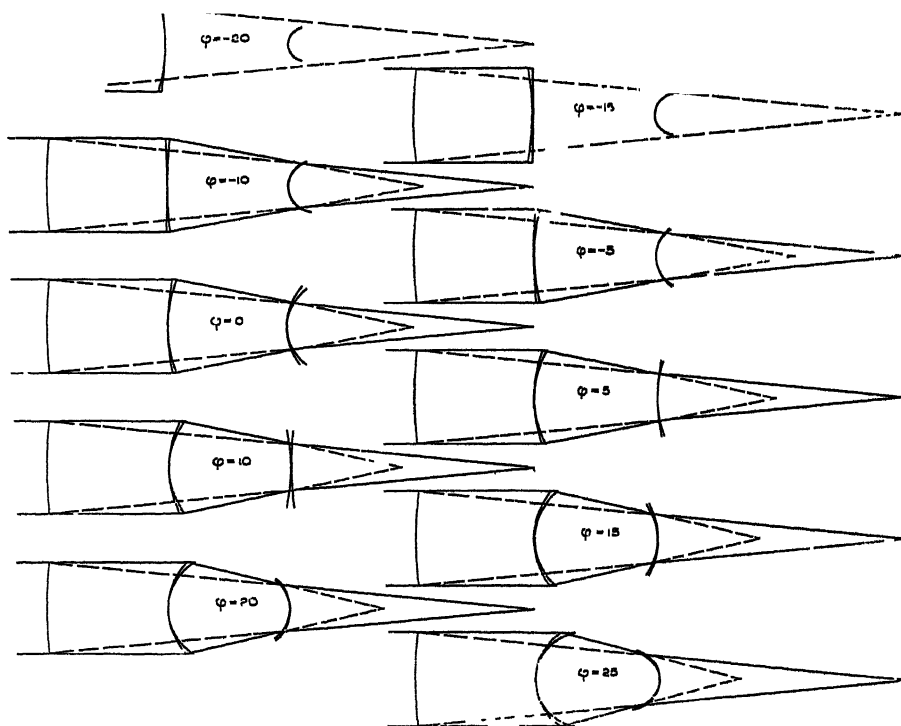


Figure 4 Variation of unit surfaces of component lenses with ϕ .

to the paraxial focus of the first lens. After travelling about half-way to this point, these rays meet the first unit surface of the negative lens, and thence proceed from the corresponding points of the second unit surface to the focus of the

complete lens. (As the second lens is negative the unit surfaces appear in their natural order.) If now the emergent rays are produced backwards to meet the incident rays, the intersections will be found to lie on the sphere, centred at the focus, of radius equal to the focal length. This is the second unit surface of the complete instrument, and is represented by the circular arc on the extreme left.

In the diagrams drawn for other values of ϕ the course of the rays may be determined by finding the intersections of the incident rays with the first unit surface of the positive lens, and those of the emergent rays with the second unit surface of the negative lens, and joining the corresponding points of the remaining unit surfaces to determine the path of the rays between the two lenses. The figure shows clearly the character of the changes in spherical aberration of the component lenses as ϕ is altered

For values of ϕ less than about 5 the unit surfaces of the two lenses tend to turn away from one another, and for values greater than 5 towards one another. For extreme values of ϕ , e. g. -20 , the curvature of the second lens is too great to allow the rays indicated to be transmitted. This is substantially the case also with $\phi = -15$ and $\phi = 25$. Systems that are to work at a fairly large aperture (the rays are drawn for an aperture of $f/5$ for the complete lens) must correspond to values of ϕ well within these limits. A value of about 5 would be very satisfactory. (For freedom from distortion the value would be 16 with $\omega = 0.75$.) To find whether such a value is consistent with the use of spherical surfaces alone necessitates more detailed consideration

§ 5 NUMERICAL INVESTIGATION

For the purpose of investigating the matter numerically it is necessary to take into account the magnitude of the functions of the refractive indices which have already been mentioned. No advantage will be gained at this stage by basing the calculations on the properties of glasses given in manufacturers' catalogues. Instead, two refractive indices are chosen as is thought convenient, and N is varied. The results which follow are for a crown glass with $\omega_1 = 0.64$ and a flint with $\omega_2 = 0.60$, where ω is the reciprocal of the refractive index. It is much more convenient in dealing with aberrations to take the reciprocal than the refractive index itself.

The calculation of the surface curvatures can be carried out in any of the familiar ways. In the method used at the National Physical Laboratory they are obtained incidentally in the calculation of the six aberrational coefficients. These curvatures for the unbent forms of doublet (with crown glass leading) and for both kinds of symmetrical triplet are shown in figure 5. The lines marked "First" and "Last" represent the curvatures of the external surfaces in all these lenses. That marked D gives the curvature of the cemented surface of the doublet. The diagram shows that it is important to discover a system which gives a fairly small value of N —a value numerically less than 5 is very desirable. The lines marked " $2(\gamma_1)$ " and " $3(\gamma_1)$ " give the curvatures of the cemented surfaces when the crown lens is divided, and similarly those marked " $2(\gamma_2)$ " and " $3(\gamma_2)$ " the curvatures when the flint lens is divided. As would be expected, these curvatures are much less than that of the cemented surface of a doublet, particularly when the positive component is divided. It does not

necessarily follow that this advantage is retained in the final form of the lens, for the triple lens may need to be bent more than the doublet. Any bending is represented on this diagram by a change in the position of the zero of the curvature scale.

Figure 6 shows how the four coefficients required to determine the properties of cemented doublets depend on the value of N . The changes in γ are much more important than those of the other functions. It is mainly the value of γ which determines the suitability of a combination of glasses for any given purpose. This value of γ is to be compared with the value which is required to correct the system. The equations to be used are, for the first component,

$$\beta + \rho(1 + \omega) = 2 + \omega + \frac{1}{2}\phi,$$

$$\gamma + 2\rho\tilde{\gamma} + \rho^2(1 + 2\omega) = 5 + 2\omega + 2\phi,$$

and for the second component

$$\beta + \rho(1 + \omega) = -3(2 + \omega) + \phi,$$

$$\gamma + 2\rho\tilde{\gamma} + \rho^2(1 + 2\omega) = 9(5 + 2\omega) + 4\phi.$$

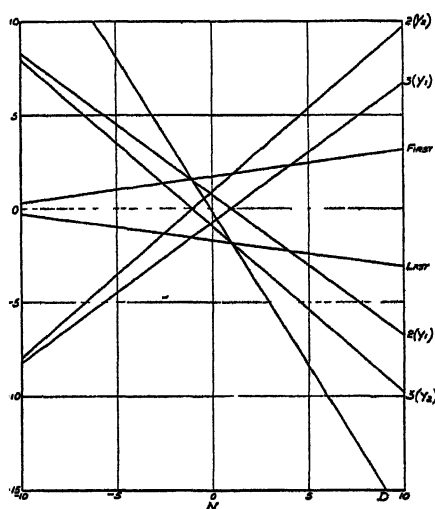


Figure 5 Curvatures of unbent lens surfaces

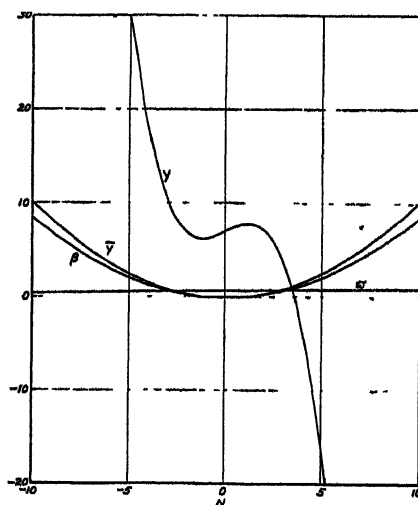


Figure 6. Aberrational coefficients for cemented doublets.

The first equation of each pair is used to determine the value of ρ , and the desirable value of γ is found by substituting for ρ in the second equation. As they stand, the equations apply to components with the crown lens leading; when the flint lens leads the signs of β and $\tilde{\gamma}$ are to be changed.

Figure 7 shows the curves giving the desirable values of γ for several values of ϕ ; the cubic curve giving the value of γ for these glasses when all the surfaces are spherical is repeated from figure 6. It will be seen that the value of N where the last curve intersects the others extends from about $N=3$ upwards. The minimum corresponds to a value of ϕ of about 8. Since N is positive for all intersections, the glasses to be used are a normal crown and a normal flint. For the minimum value of N a somewhat extreme pair would be required, but good combinations seem possible for values of ϕ between, say, -6 and 24 . Figure 8

shows the corresponding curves for the positive component when the flint lens leads. The deductions to be drawn in the two cases are very similar.

The vertical displacement between any ϕ curve and the cubic curve for any value of N indicates the amount of figuring of the lens that would be needed. It is immaterial whether this is done on the first or the last surface. The scale is

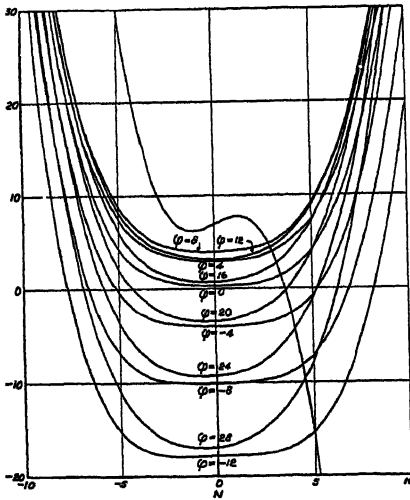


Figure 7 First component crown leading

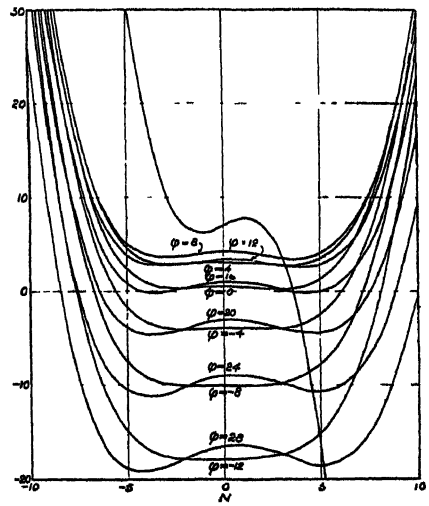


Figure 8. First component flint leading

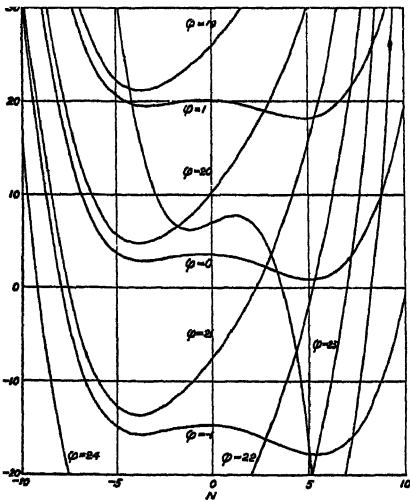


Figure 9 Second component crown leading

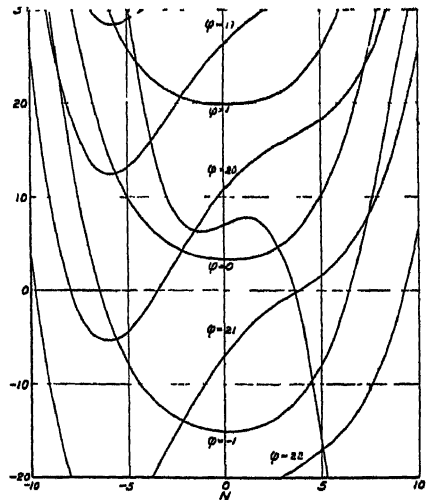


Figure 10 Second component . flint leading

such that when $N=1$ the distance of the cubic curve above $\phi=0$ represents the figuring that would be needed to make a single lens of index 1.625 aplanatic, and when $N=-1$ that required if the index were 1.66667.

Figure 9 gives results for the second component with crown lens leading. The curves are here much more widely spaced, and the minimum value of N is negative and probably numerically too large to be attractive. Positive values

of ϕ less than 20 that involve no figuring would have to be made from a pair of glasses with the greater dispersion in that of lower refractive index. For negative values of ϕ or positive values greater than 21 a normal pair of glasses would be used. If a negative value of N were accepted as a basis of construction it would probably be desirable to modify the initial assumptions to the extent of using a negative component of somewhat greater power than the positive, and thus enable a very small value for the total Petzval sum to be maintained.

Figure 10 shows the curves for the negative lens when the flint glass leads. The inferences to be drawn are very similar to those from the previous case.

Doublet lenses represent one of the extremes within which it is desirable to work. Symmetrical triplets represent the other extreme. Figure 11 shows in full lines the curves for γ_1 and γ_2 as well as for γ , and in broken lines those for

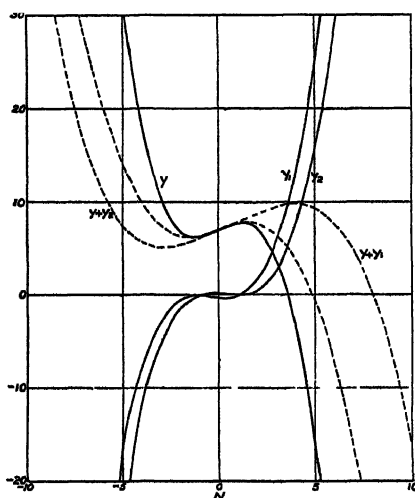


Figure 11 Aberrational coefficients for triple lenses.

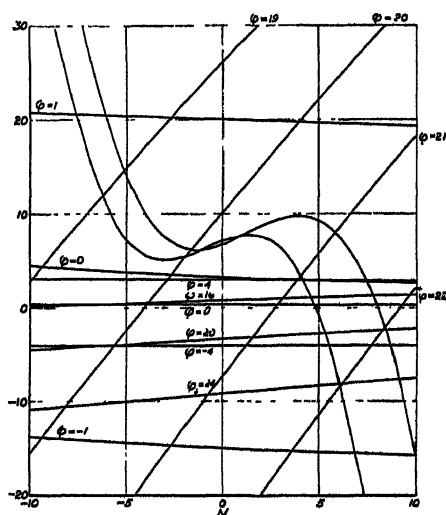


Figure 12 Symmetrical triple lenses.

$\gamma + \gamma_1$ and $\gamma + \gamma_2$. It is possible to make from given glasses a cemented lens having properties intermediate between those indicated by the appropriate doublet and triplet curves.*

Figure 12 gives curves for both forms of symmetrical triplets. Those relating to the positive component have the values of ϕ recorded in the middle of the diagram. The curves for the negative component have their ϕ values inserted near the margin of the diagram. These curves are to be considered in connexion with the $\gamma + \gamma_1$ curve copied from figure 11 if the crown lens is split, and with the

* Flexibility can often be secured by not approaching either extreme very closely. For example, it has frequently been required at the National Physical Laboratory to construct highly corrected aplanatic objectives, usually two or three in number, for any given specification. If cemented doublets were made (and cementing is practically essential in many instruments for making fine measurements) it would be necessary to obtain glasses having optical properties conforming to a fairly rigid specification, but this is no longer a requirement if glasses well away from the doublet region on the side of the triplet boundary are chosen. The neighbourhood of the doublet boundary is avoided to prevent one part of the divided crown lens from being of an inconvenient shape—it would probably be a deep meniscus for glasses near the doublet locus. Where large-scale production is involved other considerations become important.

$\gamma + \gamma_2$ curve if the flint lens is split. It is not obvious from this diagram that any marked advantages are obtainable by the use of triplet components.

In addition to the values of the six aberrational coefficients, their chromatic variations are important. These are shown in figure 13. In applications an additional factor $1/(\nu_1 - \nu_2)$ has to be applied to obtain directly comparable figures for systems made with different glasses. The outstanding points are the relative unimportance of the variations in ω , β and $\tilde{\gamma}$, and again the value of glass combinations in which N is numerically small.

From inspection of the curves of figures 7, 8, 9, 10 it appears that telephoto lenses in which the same kinds of glass are used in both components are possible, and would be fairly representative of practicable systems generally. These systems have been determined for all doublet combinations using the glasses $\omega_1 = 0.64$, $\omega_2 = 0.60$, and details are given in the accompanying table. The order of the letters C and F is that in which the crown and flint glasses are arranged

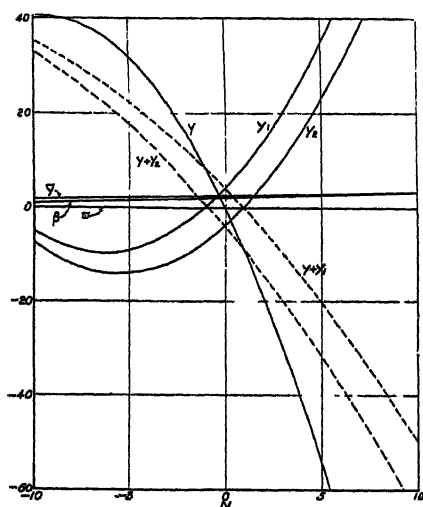


Figure 13. Chromatic variation of aberrations.

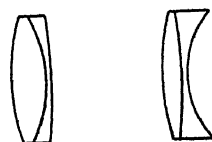


Figure 14 Preferred form of lens.

in the complete system. The first point to be observed is that the values of ϕ are very restricted. Effectively there are only two choices—either ϕ is nearly zero or it is about 21. As figure 4 indicated, the former value is the more promising. The table below shows that the smaller value of ϕ carries with it a smaller value of N , which is also a favourable indication. From the curvatures of the refracting surfaces of the eight possible systems it will be seen that the one with the least steep surfaces has a small negative value of ϕ and is arranged with the two crown glasses facing outwards. The next best system has a large ϕ , and both crown lenses precede the flint lenses to which they are cemented. A close third is the system with small ϕ and the flint components preceding the crowns. The form of construction which, on this analysis, appears the most favourable is illustrated in figure 14.

Having selected a type of lens for fuller examination, it is of interest to know whether the construction will be very sensitive to changes in the refractive index of the glass. To ascertain this it is sufficient to vary one glass only, and the crown

Table
Two-times telephoto lenses of unit focal length
Glasses $\omega_C=0.64$, $\omega_F=0.60$

	Glass order			
	CFCF	CFFC	FCCF	FCFC
N	3.407	3.434	3.580	3.607
ϕ	-0.019	-0.247	-0.083	-0.338
R_1	3.160	3.120	4.317	4.292
R_2	-4.675	-4.762	8.187	8.202
R_3	-1.064	-1.111	0.044	0.012
R_4	3.217	2.257	3.288	2.222
R_5	11.058	-1.394	11.430	-1.688
R_6	7.442	6.488	7.560	6.502
N	3.802	3.749	4.358	4.274
ϕ	21.518	21.053	22.024	21.377
R_1	6.245	6.192	7.955	7.824
R_2	-2.288	-2.250	12.905	12.735
R_3	1.912	1.873	3.485	3.359
R_4	-8.133	-9.028	-9.472	-10.880
R_5	0.400	-13.151	-0.050	-15.791
R_6	-3.800	-4.709	-5.000	-6.415

The R s denote the curvatures of the lens surfaces.

is an obvious choice. Figure 15 shows the curvatures of the six surfaces as the reciprocal crown index varies from 0.625 to 0.68: the corresponding index limits are 1.6 and 1.47. It is clear that the stability of this type of lens is satisfactory.

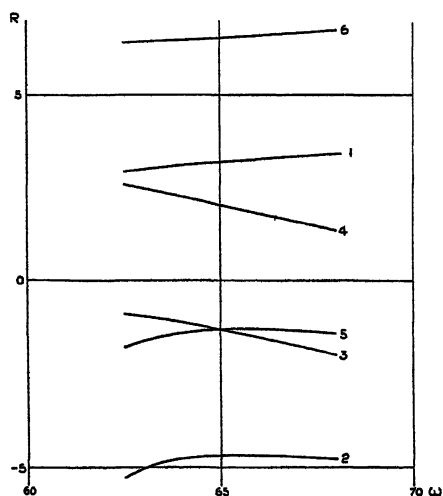


Figure 15. Changes in curvatures of CF FC lens as crown glass is altered.

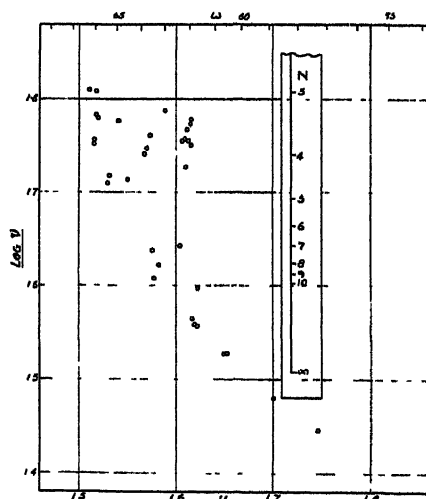


Figure 16. Glass chart and N scale

One other point calls for examination before we can say that the preliminary investigation is complete. We must know whether pairs of glasses having the relation indicated by the numerical work are within the range of available types. Figure 16 shows a chart of the standard glasses made by Messrs. Chance Bros. with the ν 's plotted logarithmically. A scale of values of ω has been added at the top of the diagram. Placed on it is a scale showing the values of N . The infinity mark on this scale is placed on the ν level corresponding to one of the glasses—in this case a flint of reciprocal index 0.60 (in the absence of a glass the corresponding point on the line of dense flints has been taken), and the value of N corresponding to any crown glass used in combination with this flint is then read off the scale. By comparing this chart with the values recorded in the table it is clear that suitable glasses are possible. The results of calculations on the preferred form of lens with different indices shows that a good choice of glasses is possible for constructing lenses of this type.

§ 6. SUBSEQUENT CALCULATIONS

So far as general exploration is concerned this investigation is now substantially completed. The next stage may either be the determination of the aberrational coefficients of the next order, still with undetermined glasses, or, as is more usual, the selection of probable glasses for a trial instrument, the insertion of suitable thicknesses in the lenses, and the introduction of small changes in the curvatures to take account of the effect of glass thickness. Experience with cemented lenses roughly comparable with those of the system portrayed in figure 14 leads one to expect that, at least so far as the central portion of the field is concerned, no particular difficulties will arise in making these alterations. This subsequent work is not considered here, for the purpose of this paper is to illustrate a method of optical prospecting developed at the National Physical Laboratory.

§ 7. ACKNOWLEDGMENT

The work described above was carried out in the Light Division of the National Physical Laboratory on behalf of the Ministry of Aircraft Production, by whose permission this paper is published.

APPENDIX

The formulae used in calculating the six aberrational coefficients are as follows :

$$a_1 = \frac{1}{2}(1 + N), \quad a_2 = \frac{1}{2}(1 - N), \quad a = a_1 + a_2 = 1,$$

$$j_1 = \frac{a_1}{1 - \omega_1}, \quad j_2 = \frac{a_2}{1 - \omega_2}, \quad j = j_1 + j_2,$$

$$k_1 = j_1 \omega_1 = j_1 - a_1, \quad k_2 = j_2 \omega_2 = j_2 - a_2, \quad k = k_1 + k_2.$$

k , k_1 and k_2 are the total curvatures of the achromatic lens and of its crown and flint components respectively.

$$w = a_1 \omega_1 + a_2 \omega_2, \quad \alpha = -j_1 j_2 (\omega_1 - \omega_2), \quad \beta = \alpha (\omega_1 + \omega_2), \quad \tilde{\gamma} = 2\beta - \alpha,$$

$$\gamma_1 = 2\beta j_1 - \alpha k_1, \quad \gamma_2 = -2\beta j_2 + \alpha k_2, \quad \gamma + \gamma_1 + \gamma_2 = j^2 + \alpha k_1 - \alpha k_2.$$

When ρ is eliminated between the equations for coma and spherical aberration, the quantity $\alpha(1+\varpi)-\beta$ occurs. An independent check on part of the calculation is obtained by seeing that $-a_1a_2j(\omega_1-\omega_2)$ has the same value. The standard curvatures of the doublet are $\frac{1}{2}k$, $-\frac{1}{2}(k_1-k_2)$, $-\frac{1}{2}k$, and those of a divided triplet (say crown divided) are $\frac{1}{2}k$, $\frac{1}{2}(k_2-\theta k_1)$, $-\frac{1}{2}(k_2+\theta k_1)$, $-\frac{1}{2}k$. The time required for calculating these six quantities and the standard curvatures is between one and five minutes.

The formulae for the chromatic variations are

$$\begin{aligned}\Delta\varpi &= \omega_1^2 - \omega_2^2, & \Delta\alpha &= k, & \Delta\beta &= j(\omega_1^2 + \omega_2^2), & \Delta\tilde{\gamma} &= 2\Delta\beta - \Delta\alpha, \\ \Delta\gamma_1 &= 2j^2\omega_2^2 + k(k_1 - 2k_2), & \Delta\gamma_2 &= -2j^2\omega_1^2 + k(2k_1 - k_2), \\ \Delta(\gamma + \gamma_1 + \gamma_2) &= k(k_1 - k_2).\end{aligned}$$

They follow immediately from the principal formulae on putting

$$\Delta a_1 = \Delta j_1 = -\Delta a_2 = -\Delta j_2 = \epsilon$$

and taking the coefficients of ϵ .

DISCUSSION

MR G. A. RICHMOND Mr Smith's paper is certainly very interesting and instructive, and should be welcomed by lens designers as the basis for further investigation.

There are one or two remarks I should like to submit, but as I have not had any opportunity of doing any computing on Mr Smith's preferred form, they must be treated as my own personal assumptions. From the construction of the lens system, it will be agreed that as far as central definition is concerned, a high degree of perfection appears to be obtainable, but it would be most interesting to have the full story of the off-axis aberrations for comparison with existing types of telephoto lenses. If the definition of Mr Smith's lens at the edge of the field is no better than that of other types having less distortion, then I think his form will be somewhat at a disadvantage.

The condition of few glass-to-air surfaces is an advantage from the point of view of contrast, but I am not sure that it is advisable to cement components of such large diameter. After all, the adjacent surfaces can be well made, and the extra variables might enable the designer to obtain better general definition.

For reasons previously mentioned, I can only express the doubt regarding higher-order aberrations at the margin of the field, and if this doubt is justified—and I sincerely hope it is not—I shall feel that Mr. Smith has tied his hands unnecessarily by correcting each component in itself as well as cementing them.

In conclusion, I would point out that while it is recognized that distortion in long-focus reconnaissance lenses is of little consequence, cases may arise in which it must be considered.

AUTHOR'S reply. The paper should be regarded as an illustration of one method of reaching a provisional design. In the earliest stages, when the number of degrees of freedom is potentially large, it seems reasonable to reduce these by assuming the very simple forms of construction are advantageous, and to find, under these conditions, what can be done to remove low-order aberrations. Having found any types of construction promising enough for fuller investigation, an increased number of variables will probably be required to control higher-order aberrations or to meet other conditions. There is nothing in the method described in this paper to prevent this. For example, the conditions involving ϕ as a free variable are not affected if the juxtaposed crown and flint surfaces differ in curvature, or if more than two components are used, or non-spherical surfaces are employed. Changes of this kind may be freely introduced and the new parameters given values which enable conditions involving both the first- and the second-order aberrations to be met. The region explored in the second stage will normally be limited in view of the conclusions reached in the first.

I agree with Mr Richmond that in this particular case improvement is likely to be attained by increasing the number of glass-air surfaces. Some troubles will be sure to arise from refraction at the cemented surface of the positive component, and in practice, when a large relative aperture is required, at least four glass-air surfaces are likely to be used. While mentioning cemented surfaces, let me say that I realize the practical difficulties experienced in cementing large lenses successfully, and that these are increased by the extreme temperatures to which the lenses may be subjected. Nevertheless, cemented combinations have considerable advantages, and I hope ways of cementing lenses of large diameter will be thoroughly investigated in the future.

I am glad to have Mr Richmond's view of the defects likely to be found in a lens of what I have called the preferred form. At the moment nothing more has been done than to ascertain that, while considerable higher-order aberrations are present, they are not of such a magnitude as to discourage the attempt to design a system based on this form. I hope it will be possible to make the comparison Mr Richmond suggests between this and the customary forms of construction of telephoto lenses.

VARIATIONAL FORMULÆ IN OPTICS

BY T. SMITH, F.R.S.,
Teddington

MS received 8 June 1945

ABSTRACT. Formulae for calculating the effect in a complete optical instrument of small changes in the powers and separations of its component parts are collected together. The coefficients in these formulae need not be specially calculated, for they are already known if rays have been traced through the original system algebraically.

§1. INTRODUCTION

IN recent papers Prof. M'Aulay and Mr. Cruickshank (1945) have given formulae for calculating the effect of small changes in optical systems, and have described their method of using them in working out the constructional specifications of optical instruments required by the Australian Armed Forces. They have based their work on Conrady's well-known book, and the fact that they have adopted his notation will no doubt be much appreciated by the many designers who use it. It is interesting that these authors, with no previous experience of optical designing, and presumably with no access to other works on technical optics, have realized the importance of branching out in this way. They have made useful additions to the normal computing routine, and it is rather surprising that formulae of this kind have been used so little, if at all, by many workers who rely on trigonometrical methods.

Expressions for the same purposes, constructed for use in conjunction with algebraic methods of lens calculation, have been in use for many years. To the writer a completely algebraic method seems to be altogether preferable to a trigonometric alternative. For the purposes now considered it yields the formulae desired more simply, and in application involves less calculation. Most of the formulae employed appeared originally in papers dealing with other problems, and it will probably be useful to collect them here. The notation used has been found convenient in many algebraic investigations. Conrady's and other trigonometric notations are not suitable.

Algebraic formulae for lens calculations can be constructed in very great variety. Those of a general character have common structural features which may not always be obvious. The formulae to be given depend on this structure and have a wide range of application. No difficulties are involved in the special applications, and the formulae are best developed in a general form. The essential property is that the characteristics of a symmetrical optical instrument can be expressed in terms of four functions which are regarded as the elements of a square matrix of order 2. When several lenses are arranged to form an instrument, the four functions of the combination are found from those of the component lenses by multiplying the corresponding matrices in their proper order. The determinant corresponding to each matrix has the value unity, so that only three of the functions are independent.

§2 MATRIX MULTIPLICATION

A brief explanation of matrix multiplication should perhaps be included for the benefit of some readers. A matrix is a collection of elements which are arranged in rows and columns and are identified by their positions. If $m_{p,q}$ is the element which lies at the intersection of the p th row and q th column of a matrix m , and $n_{q,r}$ is similarly the element which lies at the intersection of the q th row and r th column of a matrix n , the element of the product mn of the two matrices which lies in the p th row and r th column is by definition $\sum_q m_{p,q} n_{q,r}$, where the summation extends over all the columns of the factor m on the left and all the rows of the factor n on the right. (The number of rows in n is assumed to be equal to the number of columns in m .) It follows from this definition that in general the products mn and nm are different, so that matrices do not obey the commutative law for multiplication. With this exception they obey the elementary laws which hold in scalar algebra. For example, in forming the product of three matrices u, v, w , we may either form the product uv first and post-multiply it by w , or first form the product vw and pre-multiply it by u .

The matrices to be considered here have two rows and two columns, and the standard notation employed is

$$m = \begin{pmatrix} B & D \\ A & C \end{pmatrix}.$$

From the law of multiplication, if $m_1 m_2 = m_{1,2}$,

$$\begin{aligned} A_{1,2} &= A_1 B_2 + C_1 A_2, & C_{1,2} &= A_1 D_2 + C_1 C_2, \\ B_{1,2} &= B_1 B_2 + D_1 A_2, & D_{1,2} &= B_1 D_2 + D_1 C_2 \end{aligned}$$

This law of multiplication is identical with one of the common formulae for multiplying determinants, and as the determinant corresponding to any elementary matrix has the value unity, it follows that in all cases

$$BC - AD = 1.$$

If E denotes the matrix

$$\begin{pmatrix} 1 & . \\ . & 1 \end{pmatrix}$$

(where the dots represent zero elements), the law of multiplication gives

$$Em = m = mE,$$

so that E is the analogue of the scalar number 1 and commutes with all matrices m .

If m is pre- and post-multiplied by

$$\begin{pmatrix} C & -D \\ -A & B \end{pmatrix},$$

the product in both cases is E . This matrix therefore commutes with m and has the properties of an inverse with respect to m ; it is appropriately denoted by m^{-1} . It is easy to show that if $m_1 m_2 = m_{1,2}$, then $m_{1,2}^{-1} = m_2^{-1} m_1^{-1}$, i. e. the order of the factors in the inverse of a product is the reverse of that in the original product.

The only further point that appears to need explanation is the meaning of a matrix equation such as $m = n$. This equation implies that any element of m is equal to the correspondingly placed element of n ; it is thus equivalent to several scalar equations.

§ 3. THE MATRIX ELEMENTS

The events to be recorded in order to represent the passage of light through an optical instrument are the changes of direction brought about by reflexion or refraction at a surface, and the displacements on proceeding from one surface to the next. The instrumental matrix m is built up from matrix factors which represent alternately events of these two types. For the directional changes, the factor matrix has D zero and for displacements A is zero. In most cases both B and C are equal to unity in the elementary factors, and in all the product BC is equal to unity. Different systems can be constructed with somewhat different meanings for the four functions, but it may be assumed here that A represents the power of a surface or of an instrument, and $-D$ the distance along a ray between successive surfaces divided by the refractive index, or, for a complete instrument, D is the apparent distance between the extreme points for which the calculation is made. B and C are angular or linear magnifications. Assuming that the matrix relates to a (possibly) skew ray, the precise meanings can be seen in all cases from the equation

$$\begin{pmatrix} \xi' & -x' \\ \eta' & -y' \end{pmatrix} = \begin{pmatrix} \xi & -x \\ \eta & -y \end{pmatrix} \begin{pmatrix} B & D \\ A & C \end{pmatrix},$$

where x, y are the co-ordinates of the initial, and x', y' of the final point of the ray, measured in planes normal to the axis of symmetry, and ξ, η and ξ', η' are the products of the corresponding direction cosines and the refractive indices of the first and last media. The same A, B, C, D clearly determine the sagittal focal lines. In all the elementary factors B and C are both equal to unity.

If the matrix relates to tangential focal lines, the displacement factor is the same as before, but the elementary refraction factor is

$$\frac{\cos \phi}{\cos \phi'} \cdot \frac{A}{\cos \phi \cos \phi'} \frac{\cos \phi'}{\cos \phi}$$

in terms of the sagittal A , where ϕ and ϕ' are the angles of incidence and refraction

The significance of the functions in m in this case is shown by the equation

$$(\mu'\delta\psi' - \delta h') = (\mu\delta\psi - \delta h)m,$$

where δh is the length of the perpendicular from the initial point of the principal ray of the pencil to another ray inclined to it at the angle $\delta\psi$, and accented quantities relate to the final values. For example, if u and v are the distances of an object point and of its focal line, the relation they satisfy is

$$\begin{pmatrix} 1 & -\frac{u}{\mu} \end{pmatrix} m \mid \begin{matrix} v \\ \mu \end{matrix} \mid = 0$$

this is merely another way of saying that, when the end points are the object and image positions, $\delta h'$ depends on δh but not on $\delta\psi$, i.e. $D=0$.

§ 4. VARIATIONAL FORMULAE

The formulae to be considered express the changes in the system elements A, B, C, D due to small changes in the construction of the system, such as changes of refractive index, curvature, or axial separation. As the changes are small, their effects are directly additive, and it is only necessary to consider a single alteration to obtain the expressions required. In general, the part to be altered will be preceded by an unaltered portion of the system, and followed by another unaltered part. We therefore write

$$m = m_1 \cdot m_2 \cdot m_3,$$

where the part to be altered carries the suffix 2. On making the changes we clearly have the variational equation

$$\delta m = m_1 \cdot \delta m_2 \cdot m_3. \quad \dots\dots(1)$$

It is sometimes preferred to use equations derived from this by pre- or post-multiplication by m^{-1} , i. e. to use

$$m^{-1} \cdot \delta m = m_{2,3}^{-1} \cdot \delta m_2 \cdot m_3. \quad \dots\dots(2)$$

or

$$\delta m \cdot m^{-1} = m_1 \cdot \delta m_2 \cdot m_{1,2}^{-1}. \quad \dots\dots(3)$$

The most important cases are when the system 2 consists only of a single surface or a single separation, and the change is either an alteration in A_2 or in D_2 . For a change of power, equation (1) may be written

$$\begin{pmatrix} \delta B & \delta D \\ \delta A & \delta C \end{pmatrix} = \delta A_2 \begin{pmatrix} D_1 \\ C_1 \end{pmatrix} (B_3 \ D_3), \quad \dots\dots(4)$$

and for an altered separation

$$\begin{pmatrix} \delta B & \delta D \\ \delta A & \delta C \end{pmatrix} = \delta D_2 \begin{pmatrix} B_1 \\ A_1 \end{pmatrix} (A_3 \ C_3). \quad \dots\dots(5)$$

Thus the change δA in the power of the system due to the change of power δA_2 of the surface 2 is $C_1 B_3 \delta A_2$, and that due to the separation change δD_2 is $A_1 A_3 \delta D_2$.

It will be noted that relations derived from equation (1) involve a knowledge of the functions relating to both the preceding and following parts of the instrument; in other words, the functions require to be calculated from each end of the system. Duplicating the numerical results in this way is an excellent check on

the accuracy of the calculations, for the intermediate quantities are all different. But in many cases it will be desired to avoid this, and equations included in (2) or (3) are then appropriate. For (2) the calculated values are found from a reverse trace, and for (3) from a direct trace. In the most general case the relations are

$$A^2\delta\left(\frac{B}{A}\right) = -B_{2,3}B_3\delta A_2 + A_{2,3}B_3\delta B_2 - B_{2,3}A_3\delta C_2 + A_{2,3}A_3\delta D_2,$$

$$D^2\delta\left(\frac{C}{D}\right) = D_{2,3}D_3\delta A_2 - C_{2,3}D_3\delta B_2 + D_{2,3}C_3\delta C_2 - C_{2,3}C_3\delta D_2,$$

$$A^2\delta\left(\frac{C}{A}\right) = -C_1C_{1,2}\delta A_2 - A_1C_{1,2}\delta B_2 + C_1A_{1,2}\delta C_2 + A_1A_{1,2}\delta D_2,$$

and

$$D^2\delta\left(\frac{B}{D}\right) = D_1D_{1,2}\delta A_2 + B_1D_{1,2}\delta B_2 - D_1B_{1,2}\delta C_2 - B_1B_{1,2}\delta D_2.$$

Most frequently we require to know the changes due to the alteration of a single element. If only a separation is altered we insert zero values for δA_2 , δB_2 , δC_2 and note that $A_{1,2}=A_1$, $B_{1,2}=B_1$, $A_{2,3}=A_3$, $C_{2,3}=C_3$. For forward traces the formulae are then

$$A^2\delta\left(\frac{C}{A}\right) = -\delta\left(\frac{C}{D}\right) + 2C\delta\left(\frac{1}{D}\right) - C^2\delta\left(\frac{B}{D}\right) = A_1^2\delta D_2,$$

$$A\left\{\delta\left(\frac{1}{A}\right) - B\delta\left(\frac{C}{A}\right)\right\} = -D\left\{\delta\left(\frac{1}{D}\right) - C\delta\left(\frac{B}{D}\right)\right\} = -A_1B_1\delta D_2,$$

$$\delta\left(\frac{B}{A}\right) - 2B\delta\left(\frac{1}{A}\right) + B^2\delta\left(\frac{C}{A}\right) = -D^2\delta\left(\frac{B}{D}\right) = B_1^2\delta D_2,$$

and for backward traces are

$$A^2\delta\left(\frac{B}{A}\right) = -\delta\left(\frac{B}{D}\right) + 2B\delta\left(\frac{1}{D}\right) - B^2\delta\left(\frac{C}{D}\right) = A_3^2\delta D_2,$$

$$A\left\{\delta\left(\frac{1}{A}\right) - C\delta\left(\frac{B}{A}\right)\right\} = -D\left\{\delta\left(\frac{1}{D}\right) - B\delta\left(\frac{C}{D}\right)\right\} = -A_3C_3\delta D_2,$$

$$\delta\left(\frac{C}{A}\right) - 2C\delta\left(\frac{1}{A}\right) + C^2\delta\left(\frac{B}{A}\right) = -D^2\delta\left(\frac{C}{D}\right) = C_3^2\delta D_2.$$

Similarly if

$$m_2 = \begin{pmatrix} 1 & \\ A_2 & 1 \end{pmatrix},$$

so that only A_2 is changed, $C_{1,2}=C_1$, $D_{1,2}=D_1$, $B_{2,3}=B_3$, $D_{2,3}=D_3$ and the equations are

$$A^2\delta\left(\frac{C}{A}\right) = -C_1^2\delta A_2, \quad A\left\{\delta\left(\frac{1}{A}\right) - B\delta\left(\frac{C}{A}\right)\right\} = C_1D_1\delta A_2,$$

$$\delta\left(\frac{B}{A}\right) - 2B\delta\left(\frac{1}{A}\right) + B^2\delta\left(\frac{C}{A}\right) = D_1^2\delta A_2,$$

$$A^2\delta\left(\frac{B}{A}\right) = -B_3^2\delta A_2, \quad A\left\{\delta\left(\frac{1}{A}\right) - C\delta\left(\frac{B}{A}\right)\right\} = B_3D_3\delta A_2,$$

$$\delta\left(\frac{C}{A}\right) - 2C\delta\left(\frac{1}{A}\right) + C^2\delta\left(\frac{B}{A}\right) = -D_3^2\delta A_2,$$

where the expressions in terms of A may be replaced by their equivalents in terms of D already given. It will be noted that use has been made of equations (4) and (5) as well as of (2) and (3) to construct sets of formulae for one-way operation.

If

$$m_2 = \begin{bmatrix} \frac{\cos \phi}{\cos \phi'} \\ A_2 \quad \frac{\cos \phi'}{\cos \phi} \end{bmatrix},$$

and only A_2 is changed, the right sides of the forward formulae just given each contain the additional factor $\frac{\cos \phi'}{\cos \phi}$ and the backward formulae the factor $\frac{\cos \phi}{\cos \phi'}$.

Many applications of these relations concern the determination of changes of power or separation required to bring about desired alterations in the functions for the complete instrument rather than the converse problem so far considered. The changes are determined by the solution of a number of linear simultaneous equations, and it is useful to have some guide to aid in selecting the most suitable elements for alteration. To ensure that all changes are small, the determinant of the coefficients in the equations of the elements to be modified must not be small. As a rule functions relating to a number of rays will be involved, and the type of construction enters into the problem. But some guidance can be gained from the coefficients for a single ray. A typical case is when the focal length and back focal length, i. e. $1/A$ and C/A , are to be adjusted. We can then regard such quantities as $C_1 \delta A_2$ or $A_1 \delta D_2$ as independent variables, with C_1 and D_1 or A_1 and B_1 as coefficients. Changing for convenience to a double suffix solution, the matrix relations give for an instrument with n spaces

$$\begin{vmatrix} C_{1,p-1} & D_{1,p-1} \\ C_{1,q} & D_{1,q} \end{vmatrix} = D_{p,q} = \begin{vmatrix} B_{q+1,n} & D_{q+1,n} \\ B_{p,n} & D_{p,n} \end{vmatrix},$$

and for one with n surfaces

$$\begin{vmatrix} A_{1,p-1} & B_{1,p-1} \\ A_{1,q} & B_{1,q} \end{vmatrix} = -A_{p,q} = \begin{vmatrix} A_{q+1,n} & C_{q+1,n} \\ A_{p,n} & C_{p,n} \end{vmatrix},$$

so that in one case the apparent distance between the surfaces to be altered should be large, and in the other the power of the system between the separations to be changed should be large. If these quantities are small it will usually only be practicable to satisfy one of the selected conditions instead of both.

If three conditions are imposed instead of two, none of the three apparent distances, or none of the intervening powers, may be small. This is evident if, for instance, we note that

$$\begin{vmatrix} C_{1,p-1}^2 & C_{1,p-1} D_{1,p} & D_{1,p-1}^2 \\ C_{1,q}^2 & C_{1,q} D_{1,q} & D_{1,q}^2 \\ C_{1,r}^2 & C_{1,r} D_{1,r} & D_{1,r}^2 \end{vmatrix} = D_{p,q} D_{q+1,r} D_{p,r},$$

In his paper already mentioned, Prof. M'Aulay (1945) puts forward approximate formulae for focal displacements. It may be well to point out that I gave the exact law which applies in these cases many years ago (1922/23), and also the form of the sine condition for oblique rays (1929/30), which is given by the two equations

$$\begin{aligned} Axx' - x\xi + x'\xi' + D\xi\xi' &= 0, \\ Ayy' - y\eta + y'\eta' + D\eta\eta' &= 0. \end{aligned}$$

Since, following Conrady, this paper ascribes the relation $x\xi = x'\xi'$ to Lagrange, it should be pointed out now, when it is being repeated in these *Proceedings*, that this is one of several historical inaccuracies to be found in Conrady's book.

REFERENCES

- CRUICKSHANK, F. D., 1945. "A system of transfer coefficients for use in the design of lens systems." *Proc Phys. Soc.* 57, 350, 362, 419, 426, 430.
 M'AULAY, A. L., 1945. "A transfer method for deriving the effect on the image. . . from ray changes produced at a given surface." *Proc Phys Soc.* 57, 435.
 M'AULAY, A. L. and CRUICKSHANK, F. D., 1945. "A differential method of adjusting the aberration of a lens system." *Proc Phys Soc.* 57, 302.
 SMITH, T., 1922/3. "The optical cosine law". *Trans. Opt. Soc.* 24, 31.
 SMITH, T., 1929/30. "The general form of the Smith-Helmholtz equation." *Trans. Opt. Soc.* 31, 241.

THE ELECTRON MICROSCOPE AND ITS APPLICATIONS

BY F. W. CUCKOW,
Teddington

Lecture-Survey delivered 23 February 1945

THE speaker mentioned the various forms of electron microscope—those producing images of self-emitting objects; those producing point sources of electrons which were then used to project shadows of nearby objects; scanning microscopes in which the object was scanned by a fine electron probe and images of the object obtained in terms of the secondary emission; and finally, a type very similar in its component parts to an ordinary light microscope, the transmission electron microscope. The greater part of the lecture was devoted to this last type.

After tracing events leading to the development of electron microscopes, the resolving powers to be expected were discussed together with the limitations set by existing electron lenses.

Slides were shown in illustration of the various practical electron microscopes already constructed in America, Belgium, Britain, Canada and Germany, including the commercial instruments produced in the first and last named countries.

It was pointed out that the maximum "aperture" of existing electron microscopes was only about one-thousandth of that attained in high-power light microscopes on account of the severe spherical aberration affecting electron lenses.

Associated with this small "aperture", however, was a far greater depth of focus.

The techniques for specimen preparation and the applications of the instrument were discussed together. New information has been obtained on the fine structure of diatoms, which are found to be large enough to be mounted over the holes in a fine wire mesh.

Most biological material and specimens such as fine powders were mounted on films of nitrocellulose or other plastic mounted in turn on the fine wire mesh. The plastic films have to be of the order of one-millionth of an inch in thickness to avoid undue scattering of the electron beam. Filter-passing viruses, and powders, such as paint pigments, ceramic materials and fillers, and smoke particles were mounted in this manner.

The transmission type of electron microscope could not deal directly with massive specimens requiring surface examination, but such specimens could sometimes be examined indirectly by "replica" techniques; these had been used in studying the microstructure of teeth and metals.

Finally, the migratory bands, produced in electron micrographs by electron diffraction from the atomic planes of crystalline specimens, were mentioned, and also electron diffraction "past a straight edge" giving a banded pattern exactly comparable with that obtained in the well-known experiment with light.

DISCUSSION

Mr. C. G. WYNNE. It is well known that the resolving power of a lens may be considerably increased, at the expense of some loss of contrast, by the use of an annular aperture. This effect would appear to be particularly suitable for use in the electron microscope lenses, with their heavy spherical aberration, and I should be interested to know whether it has been tried. In the presence of uncorrected spherical aberration the obscuring of a central part of the beam without any increase in the maximum angular aperture of the beam (2α in Mr. Cuckow's lecture) would give increased resolution and a reduction in the effective aberration, together with reduced intensity, and this would not be conditional on accurate centring of the annular aperture, which it is suggested might be difficult. If this possible difficulty of centring could be overcome (for example by using the annular aperture at such a position that it could be imaged in the same way as the source itself is imaged for centring) then the improved aberration would permit a small increase of α , with a further gain in resolving power. As α is still further increased, the effective spherical aberration may be controlled by reducing the width of the annulus; since the width of a constant-area annulus decreases proportionally to the first power of the reciprocal of its increasing mean radius, while the aberration will increase as higher powers, large increases of α will necessitate a rapid decrease of aperture area and hence of intensity. It would appear that this will be the limiting factor in the process, so that it would be interesting to know to what extent the intensity of the source could be increased to allow the fullest scope for the method.

Mr. R. N. CROOKS. Members will probably be interested to hear that a number of electron microscopes are now being made in this country, and the firm concerned is the one that made the first instrument illustrated by Mr. Cuckow. Although production was suspended until recently owing to more pressing war needs, there is no reason to anticipate that the new instruments will not give as good a performance as the ones now being manufactured in the United States. The first instrument is expected to be ready about the middle of this year.

Mr. M. E. HAINE. I would like to say a few words concerning the history of the development of the electron microscope.

Early work was carried out mainly in Germany by Knoll and Ruska, these workers having constructed their first transmission-type instrument about 1931. Further work was done in the early thirties in Germany by the same workers and others, and by Marton in Brussels.

In 1935, in this country, work was commenced on the construction of an instrument to the design of Professor Martin, of the Imperial College. The instrument was constructed in the Research Department of Metropolitan-Vickers, and on completion was transferred to the Imperial College, and several years later to the N P L. The instrument never got past the stage of being an experimental model, although it ultimately gave a good performance. Further work in this country was held up by the advent of the war until quite recently, but in the late thirties Canadian workers at Toronto, Prebus and Hillier, developed a first-class instrument.

By 1938 commercial instruments were available in Germany, and shortly after, R C A of America, having taken Marton from Belgium and Hillier from Toronto on their staff, produced their model B, which is the instrument mainly described in this paper.

In the last year, work has been recommenced in this country, and a war-time "Utility" electron microscope is now on the market. This instrument was designed and is being produced by the Research Department of Metropolitan-Vickers in Manchester.

With reference to a previous speaker's remarks concerning the use of an annular aperture in the objective stage of the microscope, it should be noted that the width of such an aperture must be altered by an amount inversely proportional to the square of the radius of the annulus in order to keep the spherical aberration below the diffraction error. The area of the aperture will then be inversely proportional to the radius. This will result in a reduction in intensity, unless the aperture is placed in the condenser lens, in which case an increased current density might be utilized without over-heating the specimen. However, a further limitation now creeps in, due to the scattering of the beam in the specimen, which now produces r^3 times the aberration produced in the circular-aperture case.

MR H. H. HOPKINS. It has been suggested by Mr C. G. Wynne that the use of a narrow annular aperture stop for the electron beam would serve to reduce the effects of spherical aberration, while having advantages over the small circular stop at present used in giving increased resolution.

In the case of light waves, Rayleigh showed that the use of an annular stop gave a narrower central maximum than the usual Airy-disc distribution, together with increased intensity in the diffraction rings. He advocated the use of such a stop in solar observations, whereby the illumination could be reduced without impairing the resolving power—in fact, this is enhanced. The problem has been treated in detail by G. C. Steward.

Mr T. Smith has said that the effects of the annular aperture in the case of an off-axis object point have not been investigated. I would like to correct this impression. G. C. Steward has given a full treatment of the case, both for a spherical wave and for a wave suffering from first-order coma. Curves are drawn, and it is concluded "... this illustrates a decided gain in resolving power over the full aperture case". It would, in any case, be surprising to observe a sharp discontinuity in these effects when the object-point is moved off-axis by a distance of first-order magnitude.

The same effects are well known to microscopists in the use of so-called oblique illumination, which is, of course, a common practice.

If the electron beam had a circular stop 0.10 inch in diameter instead of the diameter of 0.02 inch which has been mentioned, the resolving power would, theoretically, be increased five times. An annular stop—if, in this case, an electron beam behaves the same as light waves—would give an even greater increase. It would, I am sure, be most interesting to know whether this is the case in practice.

REFERENCES

- RAYLEIGH, Lord, 1887/92. *Sci. Pap.* III. p. 90.
 STEWARD, G. C., 1925. "The Symmetrical Optical System", *Cambridge Tracts*, no. 25, p. 88 *et seq.*, and *Philos. Trans. A*, 225.

AUTHOR'S reply I find that most of the points raised have been dealt with by other speakers. No references have been seen to attempts to use annular apertures, and it seems to be a matter which might well repay practical attention. One possible difficulty would be the apparent necessity for accurately centring the aperture so that all rays passed through the system at a common inclination to the axis. I cannot agree with Mr Wynne that the "increased resolution" would not be conditional on accurate centring" even for such small-angle beams as are now used, for if the centre is blocked out of a non-axial beam produced by a de-centred aperture, the residual rays pass through the system at a variety of inclinations to the axis, whereas for an axial beam the residual rays pass through at a common inclination. The partly blocked out non-axial beam might, therefore, be expected to give a less good image on account of spherical aberration than the partly blocked out axial beam.

The advent of a commercially produced British instrument is awaited with much interest. Messrs Metropolitan-Vickers Electrical Company are to be congratulated on proceeding with this work under the conditions which have prevailed in this country of late.

FLAT-FIELDED SINGLET APLANATS

By C. R. BURCH, F.R.S.,

H. H. Wills Physical Laboratory, University of Bristol

MS. received 15 July 1945

ABSTRACT. The design of flat-fielded aspheric aplanatic singlets is discussed by means of the "plate diagram", and two flat-fielded singlets of $n = 1.525$ are designed. Flat-fielded mirror pairs are discussed, and it is shown that a sufficient condition for Seidel field-flatness is that object and image should lie on planes tangent to the mirror surfaces. Some formulae relating to reflecting microscope design are given.

§1 INTRODUCTION

SINGLET lenses can be designed to give flat-fielded anastigmatic images of distant objects, in the extended paraxial region, provided the refractive index exceeds a certain critical value $n = 1.602\dots$. Both surfaces must be aspherized unless $n = 1.618\dots$, when the convex face may be spherical (Burch, 1943). L. C. Martin (1944), applying Chrétien's aplanat theory to the "critical case", $n = 1.602\dots$, has obtained design formulae valid outside the extended paraxial region, and has shown by ray-tracing that the resulting lens, at $f/5$, is well corrected (as a monochromat) over a semi-field of $7^\circ 30'$. The lateral colour error of these lenses may be annulled by a diaphragm placed a short distance in front of the front surface. The longitudinal colour error of course remains. To reduce this error we may prefer to select a value of n less than $1.602\dots$, though this means sacrificing exact flat-fielded anastigmatism. So we are led to discuss the theory of flat-fielded aspheric aplanatic singlets. This may conveniently be done by the method of the plate diagram described in the previous paper (Burch, 1943).

§2. GENERAL THEORY OF APLANATIC SINGLETs

We shall use the notation of the previous paper: let the singlet have a front surface of paraxial radius ρ (positive if convex), back surface of paraxial radius R (positive concave), thickness d and refractive index n . It is necessary to define

the position of the object in some manner, and we do so by means of the position of the intermediate image, because this is referable with equal analytical simplicity to either surface. Let this image be formed in glass at a distance L inside the pole of the front surface (see figure 1). We can regard this singlet as an anastigmat to which there have been added four aspheric plates, two of these being the figurings on the ρ - and R -surfaces respectively, and the other two the negatives of the "Schmidt" plates which would be necessary at the ρ and R centres to anastigmatize the refractions at the ρ and R surfaces if these were spherical. For the more convenient summation of the effects of these aspheric plates we create an artificial *star-space* (i.e. a space in which the object is at infinity, "among the stars") by allowing a hypothetical anastigmat of unit focal length (in refractive index n) to act on the intermediate image. In this star-space we draw the *plate diagram*. Let m_1, m_2 be the strengths in star-space of the missing anastigmatizing plates of the ρ - and R -surfaces and m_3, m_4 the strengths of the ρ - and R -figurings. Let the distances in star-space between these plates be as shown in figure 1.

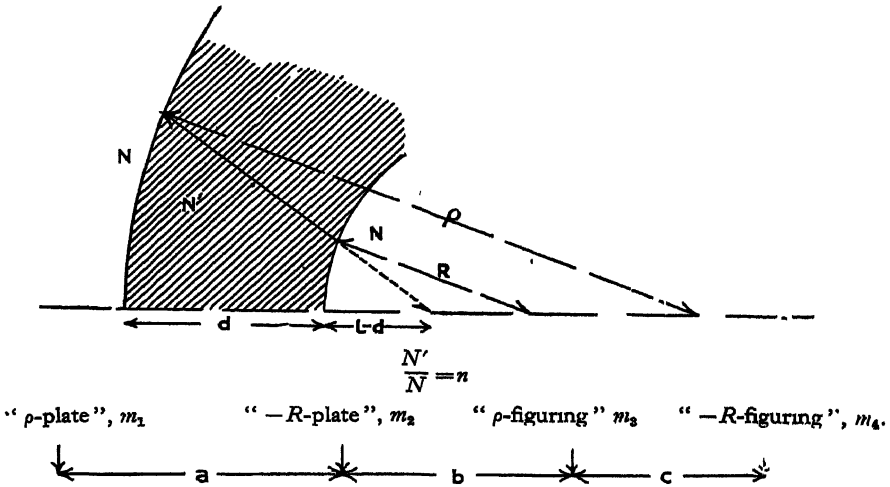


Figure 1 Two-surface system and its plate diagram

It will suffice to quote from the previous paper (Burch, 1943) the formulae for these strengths and distances.

We have then

$$\left. \begin{aligned} a &= \frac{(R+d-\rho)}{(\rho-L)(R+d-L)} & (a+b) &= \frac{\rho}{L(\rho-L)} \\ b &= \frac{(R+d)}{L(R+d-L)} & (b+c) &= \frac{R}{(L-d)(R+d-L)} \\ c &= \frac{d}{L(L-d)} & (a+b+c) &= \frac{(\rho-d)}{(L-d)(\rho-L)} \end{aligned} \right\} \dots\dots (1)$$

ρ -plate strength,

$$m_1 = \frac{H^4(n-1)n^2L^2(L-\rho)^2}{8\rho^3} \left\{ \frac{L - \left(\frac{n+1}{n}\right)\rho}{L} \right\}, \text{retardation lacking ;}$$

R -plate strength

$$m_2 = \frac{-H^4(n-1)n^2(L-d)^2(L-d-R)^2}{8R^3} \left\{ \frac{(L-d) - \left(\frac{n+1}{n}\right)R}{L-d} \right\}, \text{ retardation lacking,} \quad \dots\dots(2)$$

H being the zonal height in the artificial star-space.

Now, if in any plate diagram we equilibrate the zeroth and first moments about any pivot-point of a plate of strength m , retardation lacking, by means of two plates, one of retardation $\frac{m\alpha}{\alpha+\beta}$ distant β to the left of the $-m$ plate and one of retardation $\frac{m\beta}{\alpha+\beta}$ distant α to the right of it, we shall provide a surplus of second moment (about any pivot point) of amount $\alpha\beta$. Accordingly in aplanatizing the singlet, i.e. in annulling by suitable choice of the figurings m_3 and m_4 the zeroth and first moments of the plates m_1 and m_2 , we shall provide an excess of second moment consisting of two contributions, associable with the ρ - and R -plates respectively.

For the ρ -plate

$$\left. \begin{aligned} \alpha_\rho &= (a+b+c); \quad \beta_\rho = -(a+b); \quad [m\alpha\beta]_\rho = -(a+b)(a+b+c)m_1. \\ \text{For the } R\text{-plate} \\ \alpha_R &= (b+c); \quad \beta_R = -b; \quad [m\alpha\beta]_R = -b(b+c)m_2. \end{aligned} \right\} \dots\dots(3)$$

Thus the second moment of retardation of the aplanatic singlet is

$$\frac{-H^4(n-1)n^2}{8L(L-d)} \left[(\rho-d) \frac{L}{\rho} \left\{ \frac{L}{\rho} - \frac{n+1}{n} \right\} - (R+d) \frac{(L-d)}{R} \left\{ \frac{L-d}{R} - \frac{n+1}{n} \right\} \right]. \quad \dots\dots(4)$$

It is easy to show that plate-diagram second moments are independent of the focal length of the diagram, and are proportional to the square of the refractive index of the star-space. Eight times the coefficient of second moment of retardation of a plate diagram drawn in air, when multiplied by the square of image height, gives the negative astigmatic interfocal distance of the image. Accordingly at image height h the aplanatic single has

Negative A.I.D. =

$$\frac{-h^2(n-1)}{L(L-d)} \left[(\rho-d) \frac{L}{\rho} \left\{ \frac{L}{\rho} - \frac{n+1}{n} \right\} - (R+d) \frac{(L-d)}{R} \left\{ \frac{L-d}{R} - \frac{n+1}{n} \right\} \right]. \quad \dots\dots(5)$$

It also has Petzval depth, given by

$$\frac{h^2(n-1)}{2n} \left(\frac{1}{\rho} - \frac{1}{R} \right). \quad \dots\dots(6)$$

Let us equate (5) to κ times (6). This gives the condition that the astigmatic interfocal distance should be κ times that needed to make the surface of best definition flat, when the singlet is rendered aplanatic by figuring

$$\frac{L}{\rho} \left[\left\{ \frac{L}{\rho} - \frac{n+1}{n} \right\} (\rho-d) + \frac{\kappa(L-d)}{2n} \right] = \frac{L-d}{R} \left[\left\{ \frac{L-d}{R} - \frac{n+1}{n} \right\} (R+d) + \frac{\kappa L}{2n} \right], \quad \dots\dots(7)$$

which differs from the anastigmatism condition obtained in the previous paper only by the inclusion of the terms $\frac{\kappa(L-d)}{2n}$, $\frac{\kappa L}{2n}$.

If we fix the front radius, ρ , and the object distance, L , and satisfy (7), one of the two variables, R and d , remains at our disposal, and we may use it to annul either of the two figurings. We can achieve aplanatism with only one figured surface if, and only if, the combined first moments of the R - and ρ -plates about that surface are zero.

Thus the condition for aplanatism when the front surface only is figured is $m_1/m_2 = -a/(a+b)$, i.e.

$$\frac{L-\rho}{\rho} \left[\frac{L}{\rho} - \frac{n+1}{n} \right] = - \frac{(R+d)(R+d-L)(L-d)}{R R L} \left[\frac{L-d}{R} - \frac{n+1}{n} \right], \dots\dots (8)$$

and the expression for the negative astigmatic interfocal distance then reduces to

$$\frac{-h^2(n-1)}{\rho} \left[\frac{L}{\rho} - \frac{n+1}{n} \right] \left[\frac{R+d-\rho}{R+d-L} \right]. \dots\dots (9)$$

When the back surface only is figured, the condition for aplanatism is

$$\frac{(R+d-L)}{R} \left[\frac{L-d}{R} - \frac{n+1}{n} \right] = - \frac{(\rho-d)(L-\rho)}{\rho} \frac{L}{(L-d)} \left[\frac{L}{\rho} - \frac{n+1}{n} \right] \dots\dots (10)$$

and expression (5) reduces to

$$\frac{-h^2(n-1)}{R} \left[\frac{L-d}{R} - \frac{n+1}{n} \right] \left[\frac{R+d-\rho}{\rho-L} \right]. \dots\dots (11)$$

§ 3. SPECIAL CASE OBJECT AT INFINITY

We shall now confine our attention to the case when the object is at infinity, that is, we set $L = n\rho/(n-1)$. Let us set

$$x = d/\rho; \quad y = \frac{L-d}{R}; \quad z = R/\rho = \frac{n-(n-1)x}{(n-1)y}, \quad \dots\dots (12)$$

and the condition that the astigmatism of the aplanat should be κ times that needed to flatten the field of best definition—equation (7)—reduces to

$$\begin{aligned} & z^2 \left[x \left\{ (n+1) \frac{(n-1)^2}{n} + \kappa \frac{(n-1)}{2} + 1 \right\} - n^2 - \frac{n\kappa}{2} \right] \\ & + z \left[x^2 (n-1)^2 \frac{(2n+1)}{n} - x(n-1) \left(3n+1 + \frac{\kappa}{2} \right) + n^2 - \frac{n\kappa}{2} \right] \\ & + x[n - (n-1)x]^2 = 0. \end{aligned} \dots\dots (13)$$

When κ is fixed, equation (13) defines a curve on the (x, z) diagram passing through all the points (x, z) , representing singlet aplanats having field-flattening coefficient κ for distant objects. Two cases, $\kappa=1$ (the flat-fielded aplanats; curve E) and $\kappa=0$ (the anastigmats; curve F), are illustrated in figure 2, which has been computed for $n=1\,525$. We may conveniently show on the same diagram the two curves, A, of the aplanats which only require figuring on the front surface, and B, of those which only require figuring on the back surface.

Curve A is obtained from equation (8), which reduces to

$$(n-1)^2(1-y)(1/n+1-y)[n-(n-1)x(1-y)]=1, \quad \dots\dots(14)$$

while curve B is obtained from equation (10), which reduces to

$$y=1+\frac{1}{2n} \pm \sqrt{\frac{1}{4n^2} + \frac{1-x}{(n-1)^2+n-(n-1)x}}, \quad \dots\dots(15)$$

y being defined by (12).

The lenses of curve A have negative A.I.D. given by

$$n\rho \left(\frac{z+x-1}{z+x-\frac{n}{n-1}} \right) \quad \dots\dots(16)$$

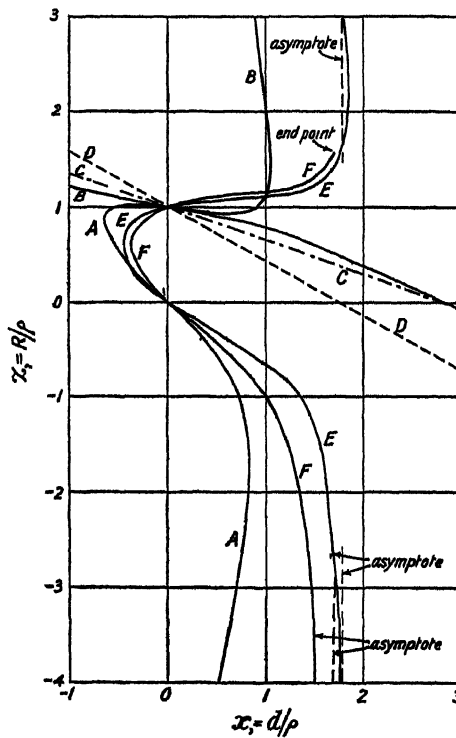


Figure 2.

from (9), while from (11) lenses of curve B have negative A I.D. of

$$+ \frac{h^2(n-1)}{z^2\rho} \left[n-(n-1)x - \left(\frac{n^2-1}{n} \right) z \right] [z+x-1] \quad \dots\dots(17)$$

For the sake of completeness we notice in the general case, when aplanatism is achieved by figuring both faces, the lens, whether flat-fielded or not, has a negative A.I.D. of amount

$$\frac{-h^2(n-1)}{n[n-(n-1)x]} \left[1-x-(z+x) \left(\frac{n-(n-1)x}{z} \right) \left(\frac{n-(n-1)x}{z} - \frac{n^2-1}{n} \right) \right]. \quad \dots\dots(18)$$

We may conveniently add to the diagram the *afocal line*, C, the equation of which is

$$z + \left(\frac{n-1}{n}\right)x = 1. \quad \dots (19)$$

The focal length of any singlet is obtained by dividing $R(=z\rho)$ by $(n-1)$ times the height above the afocal line of the representative point. Finally, we add the line D, which represents those singlets which are laterally achromatic when used as eyepieces with an objective of very great tube-length: the equation of this line is

$$z + \left(\frac{n^2-1}{n^2}\right)x = 1, \quad \dots (20)$$

and the clear distance of the eyepoint of this eyepiece in front of the ρ -surface is

$$\frac{d}{(n-1)x^2} \left[x - \frac{n^2}{n-1} \right] \quad \dots (21)$$

We are now ready to discuss the diagram. First we notice that the flat-field curve (E) intersects the anastigmat curve (F) in two places. One of these, $x=0, z=1$, is trivial, in that the lens vanishes by coincidence of the surfaces. At the second intersection, the two curves represent different lenses. Both share the properties $R/\rho=0, d/\rho=0$, and so they have the same representative point: they differ in that in the anastigmat $d/R=-1$, while in the flat-fielded lens, as we approach this point, d/R approaches $-\left(1 + \frac{1}{2n}\right)$.

ρ of course must be made infinite to obtain two physically realizable lenses. The former is the refracting analogue of the Schmidt camera: the latter is the refracting analogue of the "short" (Schmidt-Wright) system, in which enough astigmatism is introduced to annul the Petzval term of the mirror.

Next we observe that all the flat-fielded aplanats lie within the range $-0.45 < d/\rho < 1.8$ approx, and all the anastigmats within the range $-0.36 < d/\rho < 1.6$. (The anastigmat curve ends in a "singular point" at $R/\rho = 1.563 \dots$) Save in the trivial case of evanescence, none of these lenses fulfils the lateral achromatism condition of the "simple" eyepiece, and none is telescopic. Two flat-fielded aplanats exist which only need one surface figuring: one, figured on the front surface, corresponds to the intersection of curve A with the flat-field curve, having $d/\rho = -0.41, R/\rho = 0.435$. To make this lens physically realizable, we must make ρ negative: it is thus a converging meniscus, presenting a concave face to the incident light. If we set $\rho = -1$, the focal length, F , is 2.559, and the astigmatic inter-focal distance is $0.222 \dots h^2 = 0.568 \dots h^2/F$, i.e. 0.568 times that of a centrally diaphragmed thin lens of the same focal length.

The second example, in which only the back surface is figured, corresponds to the intersection of curve B with the flat-field curve. It has $d/\rho = 0.98, R = 1.10$. If we set $\rho = +1, F = 4.75$, and the astigmatic interfocal distance is $0.017 h^2, = 0.081 h^2/F$. The shape of this lens does not differ much from that of the $n = 1.618$ flat-fielded anastigmat described in the previous paper. If we "diaphragm" it at the pole of the back face, the first refraction will be nearly distortionless, while

the second will provide the rather heavy pincushion distortion of fractional amount $\frac{1}{2}(n^2-1)\theta^2$ due to refraction from glass to air. The alternative design given above, supposed diaphragmed at the pole of the front face, will give the barrel distortion of fractional amount $\frac{1}{2}\left(\frac{n^2-1}{n^2}\right)\theta^2$ due to refraction from air to glass, since the distortion due to the second refraction will obviously be small. Both distortions are increased by shifting the stop to the point giving paraxial freedom from transverse chromatism, which lies a distance D' in front of the pole of the front surface, given by

$$D' = \frac{\rho d[(n-1)R - n\rho + (n-1)d]}{[n\rho - (n-1)d]^2 - R[n^2\rho - (n^2-1)d]} \quad \dots\dots (22)$$

(This expression was derived in the previous paper.)

Thus the $\rho=1$, $d=0.98$ example has its lateral achromatizing point $0.34\dots\rho$ in front of the front surface, while the $\rho=-1$, $d=0.41$ example has it -0.315 in front of the front surface in object space—i. e. about 0.1 behind the back face in image space. When the lenses are “diaphragmed” at these points the respective fractional pincushion distortions at object angle θ are $0.67\theta^2$ and $-0.31\theta^2$.

It remains to calculate the strengths of the figurings needed in the two cases : in the artificial star-space the retardation required is m_1+m_2 (expression (2)). Transference to “fresh air” (i. e. air in front of the front face of the lens), since it

increases the zonal height by a factor $\frac{n\rho}{n-1}$, multiplies plate strengths by $\left(\frac{n-1}{n\rho}\right)^4$. This gives for the figuring needed on the front face of the first ($\rho=-1$)

example a glass thickness of $3.27h^4$ at zonal height h . Evaluating the corresponding quantity for the second example ($\rho=1$) and multiplying by $\left(\frac{L}{L-d}\right)^4$ to allow for the reduction of zonal height on passing to the back face, we find that the thickness required at height h on the back face is $0.277h^4$. This completes the extended paraxial data for these designs.

§ 4 DISTORTION OF SINGLET APLANATS

I shall give, for the sake of completeness, expressions by which the distortion of these singlet aplanats can be computed. We consider, then, the plate diagram shown in figure 1 (in glass). It is convenient to divide the distortion into six contributions—three associable with refraction at the first surface and that part of the total figuring which aplanatizes it, and three associable in a corresponding way with the second surface.

If the principal ray make angle ω_s with the axis in a star-space of index n_s , and a plate of retardation mh^4 be situated at distance S down-light from the stop, the fractional pincushion distortion due to the plate is $\omega_s^2 24mS^3/n_s$. The aplanatizing figurings $\frac{m\alpha_\rho}{\alpha_\rho+\beta_\rho}$, $\frac{m\beta_\rho}{\alpha_\rho+\beta_\rho}$ defined in § 2 have together a third down-light

moment about the ρ -centre of amount $m\alpha_\rho\beta_\rho(\alpha_\rho-\beta_\rho)$. So the fact that the

refractions are aplanatized only, and not anastigmatized, provides, when the stop is located at the refracting centre, a pincushion distortion coefficient

$\frac{4m\alpha_p\beta_p}{n_s} [\alpha_p - \beta_p]$, and correspondingly for the R -surface (I retain the subscript

s in n although for the glass star-space $n_s = n$, as a reminder of the origin of this factor.) Shifting the stop a distance S up-light increases the third moment of an aplanatic plate triad by $3S$, times its second moment. But the second moment was shown to be $m\alpha\beta$. Hence the *plate contribution* to the distortion is

$\frac{4m\alpha\beta}{n_s} \cdot [\alpha - \beta + 3S_s]$ with subscript p and R in turn. The remaining terms are

the Petzval contributions which would operate if each refraction were anastigmatized by a plate at the refracting centre. If we set $S_s = S_{p_s}$ for the case of the p -centre, and $S_{R_s} = S_{p_s} + a$ for the R -centre, the Petzval contribution is

$$-\left(\frac{n-1}{2n}\right)n_s \left[\frac{S_{p_s}}{\rho} - \frac{S_{R_s}}{R} \right] \omega_s^2 \quad \dots \dots (23)$$

We find from (1)

$$a = \frac{R+d-\rho}{(\rho-L)(R+d-L)}; \quad (\alpha-\beta)_p = \frac{(L-d)\rho + (\rho-d)L}{L(L-d)(\rho-L)};$$

$$(\alpha-\beta)_R = \frac{LR + (L-d)(R+d)}{L(L-d)(R+d-L)}, \quad \dots \dots (24)$$

while $(m\alpha\beta)_p$ and $(m\alpha\beta)_R$ are the respective coefficients of H^4 in (4), so that the *plate contributions* are

$$\frac{-(n-1)n_s}{2L(L-d)} \left[(\rho-d) \frac{L}{\rho} \left(\frac{L}{\rho} - \frac{n+1}{n} \right) \right] \left[\frac{(L-d)\rho + (\rho-d)L}{L(L-d)(\rho-L)} + 3S_{p_s} \right] \omega_s^2$$

and

$$\frac{(n-1)n}{2L(L-d)} \left[(R+d) \frac{(L-d)}{R} \left(\frac{L-d}{R} - \frac{n+1}{n} \right) \right] \left[\frac{LR + (L-d)(R+d)}{L(L-d)(R+d-L)} + 3S_{R_s} \right] \omega_s^2. \quad \dots \dots (25)$$

These formulae relate to the star-space of index $n_s = n$, defined in § 2. If we wish to transfer to a star-space of index 1, of the same focal length, we must delete the factor n_s and substitute for the brackets containing S_{p_s} , S_{R_s} the values, n_s times smaller, of the corresponding distances in the new star-space. If we transfer to a star-space of q times greater focal-length, the corresponding distances will be increased q^2 times. When the object is at infinity it is natural to select as star-space "fresh air" in front of the lens. In that case put $n_s = 1$ in the above; for a ,

$[\alpha - \beta]_p$ and $[\alpha - \beta]_R$, substitute $\frac{L^2 a}{n}$; $\frac{L^2}{n} [\alpha - \beta]$, $\frac{L^2}{n} [\alpha - \beta]_R$; and for S_{p_s} , S_{R_s}

use the actual "fresh air" distances

The distortion values given in § 3 were obtained in this way. But I have failed to find any method of simplifying (23) to (25), so the search for a distortionless flat-fielded singlet aplanat remains a problem of trial computations.

§ 3 REFLECTING FLAT-FIELDED MIRROR PAIRS

It is often salutary and sometimes instructive to set $n = -1$ in thick-lens formulae.

Let us set $\kappa = 1$, $n = -1$ in (7); then the thick singlet becomes a mirror pair and the condition for flat-fieldedness reduces to

$$\frac{L}{\rho} \left[\frac{L}{\rho} (\rho - d) - \frac{L - d}{2} \right] = \frac{(L - d)}{R} \left[\frac{(L - d)}{R} (R + d) - \frac{L}{2} \right] \quad \dots (26)$$

$$\text{or} \quad \left(\frac{L}{L - d} \right)^2 + \left(\frac{1}{2R} - \frac{1}{2\rho} \right) \frac{\rho^2}{(\rho - d)} \left(\frac{L}{L - d} \right) = \frac{\rho^2 (R + d)}{R^2 (\rho - d)} \quad \dots (27)$$

Let u be the distance of the object from the ρ -mirror, then

$$\frac{L - d}{L} = \frac{u(\rho - d) + d(\rho - u)}{u\rho} \quad \dots (28)$$

and

$$[u(\rho - d) + d(\rho - u)]^2 + \frac{(R - \rho)}{2(R + d)} \cdot Ru [u(\rho - d) + (\rho - u)] = \frac{R^2 u^2 (\rho - d)}{R + d} \quad \dots (29)$$

The corresponding conditions for anastigmatism instead of flat-fieldedness are obtained if in (27) or (29) the second term on the L.H.S.—the “Petzval”—is deleted.

Two object-positions are of special interest—at infinity, or on the surface of one of the mirrors. We shall first set $u = \infty$ when (29) reduces to

$$(\rho - 2d)^2 = \frac{R\rho[R + \rho - 2d]}{2(R + d)}$$

or

$$R = \frac{\rho(-2d)}{2\rho} [\rho - 4d \pm \sqrt{\rho^2 + 16d^2}]. \quad \dots (30)$$

It may be verified that this is identical with the condition given in another notation by Schwarzschild for extended paraxial flatness of the surface of best definition of an aplanatic two-mirror telescope objective. Such objectives have been very fully discussed by Schwarzschild, Chrétien and others. I shall therefore touch only on two points, about which I am frequently asked: Is there a design of two-mirror telescope O.G. in which (a) only the primary mirror, (b) only the secondary mirror should be aspherized? The answer to both questions is yes, but the images in both cases are inconveniently placed—unless, indeed, one is prepared to figure the one figured mirror about a centre lying outside its rim, and so make an eccentric portion of an axially symmetrical objective—to make what might be called a Herschel-Schwarzschild objective. When the primary mirror is to be figured, we must satisfy equation (14), which reduces to

$$\frac{d}{\rho} = \frac{R}{2d - \rho} \left[\frac{R + 2d - \rho}{2R + 2d - \rho} \right]^2 \quad \dots (31)$$

I think Schwarzschild himself pointed out the high “obstruction ratio” due to the secondary mirror in this case. Examples are $\rho = 1$, $d = 9/32$, $R = -7/32$ and $\rho = 1$, $d = 4/9$, $R = -1/9$.

When the secondary mirror is to be figured, we must satisfy (15), which reduces to

$$R = (\rho - 2d) \left(\frac{\rho - 2d}{\rho - d} \right) \left(1 \pm \sqrt{\frac{d}{2\lambda - \rho}} \right). \quad \dots \quad (32)$$

To make the primary mirror larger than the secondary, we must make ρ positive and $d < \rho$, while a real R demands $\rho < 2d$. To obtain a real final image we have therefore to take the negative root. The final image of this rather peculiar "Gregorian" is awkwardly placed, lying between the primary image and the secondary. Example, $\rho = 1$, $d = 0.7$, $R = -0.1722$.

On the other hand, in the reflecting microscope ("projection" O.G.) we do not object to a secondary mirror several times bigger than the primary, and so we may set ρ negative in (32), with a physically real (positive) d , and take the negative root: examples, $\rho = -1$, $d = 4$, $R = -5.4$, $\rho = -1$, $d = 7$, $R = -8.91$. If it is desired to use apertures up to 0.65 N.A.—and this I find is possible—higher order aberrations cannot be neglected. (32) then gives only a rough approximation to optimal values, and the final spacing should be determined by ray tracing or—more simply—empirically. Expression (32) was given (in another form) by Schwarzschild.

Let us now consider the other special case, when the object lies on the surface of one of the mirrors. If we set $u = d$, i.e. put the object at the surface of the R -mirror, expression (29) reduces to

$$4(\rho - d)(R + d) = R\rho, \quad \dots \quad (33)$$

the corresponding condition for anastigmatism being

$$4(\rho - d)(R + d) = R^2 \quad \dots \quad (34)$$

But (33) is a physically symmetrical expression, for it states that the product of distances from the centre of each mirror to the surface of the other is equal to the product of the focal lengths. This suggests what may be verified formally, that since the object lies on the surface of one mirror, the image will lie on the surface of the other—an elegant and unexpected result. If, in addition to flat-fieldedness, we demand anastigmatism, we must set $\rho = R$, giving $3\rho^2 = 4d^2$, and the magnification of the pair is $2 + \sqrt{3}$.

REFERENCES

- BURCH, C. R., 1943. *Proc. Phys. Soc.* 55, 433
 MARTIN, L. C., 1944. *Proc. Phys. Soc.* 56, 104

DEMONSTRATION: A FRANCK AND HERTZ CRITICAL POTENTIAL EXPERIMENT

BY J. H. SANDERS,
Clarendon Laboratory, Oxford

Demonstrated 8 June 1945

THE experiment to be described is a repetition of the well known Franck and Hertz method of finding the critical potentials of mercury vapour (Franck and Hertz, 1914) suitable for lecture demonstration. It makes use of the cathode-ray oscillograph, and shows the use of the instrument, not only in demonstrations of this sort, but also in obtaining rapid qualitative observations in experiments involving two independent variables.

The electrode structure of the vapour-filled bulb is shown in figure 1. The

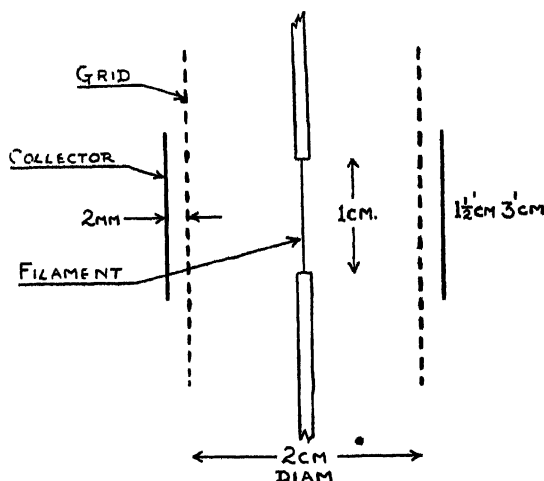


Figure 1. Electrode structure

electrodes are of nickel and the filament of 100- μ tungsten. The bulb was evacuated and the glass envelope, electrodes and filament were subjected to the usual out-gassing processes. A small quantity of mercury was distilled into the bulb, and this was then sealed off from the pumping system.

To show the curves relating collector current to grid voltage on the oscillograph screen, the oscillator shown in figure 2 is used. This delivers an alternating voltage of saw-tooth form between the filament and the grid. By using the cathode follower output, the grid swings from zero volts up to a positive voltage of about 30 volts peak with respect to the filament. The usual source of a small constant potential between the grid and collector is omitted, and the electron current flowing through the resistance in this circuit maintains the collector at a small negative potential with respect to the grid. The voltage across this

resistance is amplified and applied to the Y-plates of the oscillograph, and the saw-tooth oscillator output is applied to the X-plates.

The shape of the curves obtained has been adequately explained elsewhere (Compton and Mohler, 1924). The shape depends greatly on the pressure of the mercury vapour in the bulb. The main kinks are due to the 4.7, 4.9 and

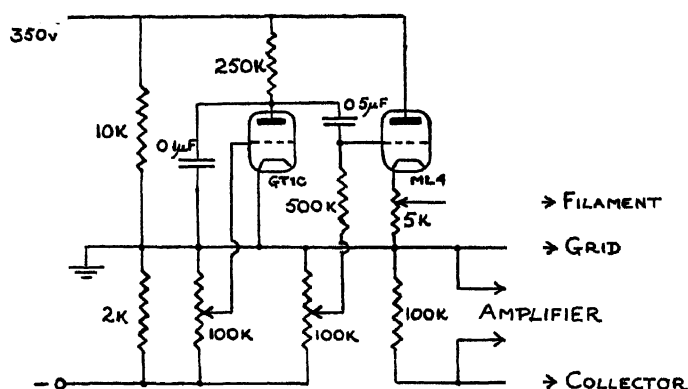


Figure 2 50-cycle saw-tooth oscillator

5.4-volt critical potentials. These three potentials are not resolved, though this has been done using a version of the Franck and Hertz apparatus (Morris, 1928). The size and number of the kinks grow as the pressure in the bulb is raised. The optimum working temperature seems to be about 120°C. A photo-



Figure 3 Curve taken at 120°C

graph taken from the oscillograph at this temperature is shown in figure 3. The first two kinks are separated by about 4.9 volts, and a smaller kink, due possibly to electrons undergoing one 4.9-volt and one 6.7-volt inelastic collision, can be seen on the side of the second main kink. The kinks appearing at higher grid voltages are due to electrons undergoing various combinations of the 4.9-volt

group of collisions and 6·7-volt collisions, as explained in the paper of Compton and Mohler.

The change of shape of the curves as the vapour pressure of the mercury in the bulb is altered may be demonstrated in a few minutes by using the oscillograph, this representing a very considerable advantage over the more conventional "D.C." approach to the experiment.

I am indebted to Mr. C. H. Collie and Dr. D. Roaf for their help in setting up this experiment.

REFERENCES

- COMPTON and MOHLER, 1924. *Bull. Nat. Res. Council, Wash.*, no. 48.
 FRANCK and HERTZ, 1914 *Verh. dtsch. Phys. Ges.* 16, 10
 MORRIS, 1928 *Phys. Rev.* 32, 447

DISCUSSION

Dr. T. B. RYMER. I should like to ask Mr. Sanders whether it is possible to use any type of commercial thyatron valve for a demonstration of this kind, in place of the specially constructed valve we have seen.

Mr. A. E. JENNINGS. Assuming that the author derived his Y deflection, which was a measure of the current passing to the outer electrode, by applying the potential drop across a high resistance in series with the small constant retarding potential to the Y-plates of the C.R.O., did this varying potential not affect the constant retarding potential, and how did this affect the validity of the experiment?

AUTHOR'S REPLY. The most important condition which has to be fulfilled in order to obtain large sharp kinks is that the drift space between the filament and the grid shall be large compared with the distance between the grid and the collecting electrode. This condition was not fulfilled by the only thyatrons which were available, and it was therefore thought better to make up a suitable valve.

Mr. Jennings is of course right in pointing out that the collector potential is not constant in this experiment. The effect of the 100 κ resistance between collector and earth will be slightly to diminish the sharpness of the kinks, as compared with the curves taken with a fixed retarding potential. As is seen in figure 3, this effect is not serious.

DEMONSTRATION : APPARATUS ILLUSTRATING THE MICHELSON STELLAR INTERFEROMETER

By P. J. TREANOR, S.J.,
 Clarendon Laboratory, Oxford

Demonstrated 8 June, 1945

THE effect upon which this instrument operates is the disappearance of the interference fringes formed by two similar apertures by the appropriate displacement of the patterns arising from neighbouring sources of light (Michelson, 1890, 1920). In a laboratory experiment the relation between the angular separation of the objects to be resolved and the separation of the diffracting apertures is most conveniently altered by varying the former, since a double slit of variable separation is rarely available. The use of a Fresnel bi-prism, and like devices for producing artificial doubles, is prohibited by the

fact that the components so produced are coherent sources, and do not therefore simulate a genuine double star

These difficulties are overcome by the use of a Wollaston quartz compound prism, which is placed (figure 1) on an optical bench in front of an illuminated

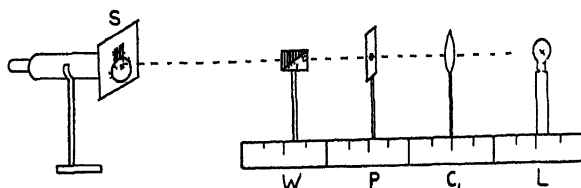


Figure 1.

pin-hole. The prism produces the components by double refraction, the light from each being polarized in mutually perpendicular directions. Since the components are selected from different planes of vibration of the natural light from the pin-hole source, they are genuinely incoherent, in the sense that they would not interfere, even if brought back into one plane by introducing a polaroid plate.

The effect is observed through a small telescope, whose objective is covered by an appropriately oriented double slit. It is convenient to cover part of the slit with a weak cylindrical lens, which has the effect of producing a greatly broadened pattern, without altering the distance between successive maxima.

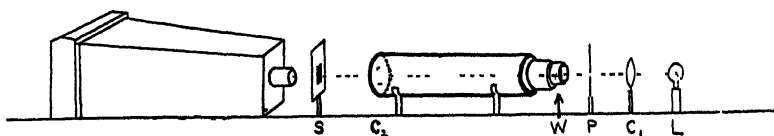


Figure 2.

The telescope should be a couple of metres from the pin-hole. Slits 2 mm. apart, of width 0.3 mm., with a cylindrical lens of 0.25 dioptres, give easily observed fringes in a telescope of 30 cm focal length, $\times 10$. Their disappearance may be observed for various separations of the components, produced by moving the Wollaston along the bench.

Figure 2 indicates how the artificial double may be used in a very similar way when it is desired to photograph this and other diffraction phenomena connected with double-star resolution. The apparatus is here adapted to the optical system of a spectrograph of 150-cm focal length, the collimating lens being the only important modification to the arrangement of figure 1. Owing to the incoherence of the light from the components and their different planes of vibration, a polaroid may be introduced between collimator and camera to vary the relative intensities of the components. I have used the diffraction photography arrangement of figure 2 to photograph the advantageous effect which a square stop affords in resolving artificial doubles of very unequal intensity.

REFERENCES

- COUDER and JACQUINOT, 1939 *C.R. Acad. Sci. Paris*, 208, 1639
 MICHELSON, 1890. *Phil. Mag.* 30, 1, 1920. *Astrophys. J.* 51, 257, 1921 *Ibid.* 53, 249.

OBITUARY NOTICES

JOHN AMBROSE FLEMING

WITH the death of Sir Ambrose Fleming the Physical Society loses the last survivor of its Founder Fellows. That bald statement does less than justice to a connection which is unique in the history of scientific societies, for on 21 March 1874, at the inaugural meeting of the Physical Society, the first communication to be presented to the Society was a paper by J. A. Fleming on the theory of the galvanic cell. During the long and distinguished career that followed this auspicious opening, Fleming contributed more than one hundred important papers to the *Proceedings* of various learned societies and institutions, and of these papers more than a third have appeared in the *Proceedings of the Physical Society*. The interval between the first and the last of these communications was no less than sixty-five years, for on 28 April 1939 Sir Ambrose Fleming read before the Society a paper on "A new method of creating electrification", illustrated by experiments. The occasion was one never to be forgotten by those who were privileged to be present. The veteran author was in fine form, and his presentation of his paper, made, as it was, in a clear and resonant voice, not read from a manuscript, logically developed without the introduction of a superfluous word, was a model for speakers not a quarter of his age. Fleming, early in his career, had obviously profited by the classic advice given to young debaters "State your case, clear your case, prove your case. sit down"; his published works bear eloquent testimony thereto.

John Ambrose Fleming was born on 29 November 1849 at Lancaster, where his father was a Congregational Minister. When he was about five years old, his parents removed to North London, where his youth was spent. Quite early in his life, he, like Clerk Maxwell before him, showed a decided inclination to investigate the "particular go" of any mechanism that came his way, and he equally early developed, with home-made apparatus, that technique of popular lecturing in which, in the years of his maturity, he was to prove himself a worthy successor to Faraday and to Tyndall.

After some years of attendance at private schools he was sent, at the age of fourteen, to University College School, then housed in a wing of University College in Gower Street. He had (in common with many other lads of similar tastes) ambitions towards the engineering profession, but the usual mode of entry thereto—a long apprenticeship and a high premium—was out of the question, so it was decided that he should work at school for the London Matriculation Examination, which he passed at the age of sixteen. Then followed two years at University College, where he studied under Carey Foster, A. W. Williamson and—an influence for life—that great teacher and erudite genius Augustus de Morgan. At eighteen he passed the first B.Sc. examination of London University, but the question of earning a living loomed large and, after a fruitless and unhappy three months in the offices of a shipbuilding firm near Dublin, he returned to London to take up work as a clerk on the Stock Exchange where, the hours of work (10 a.m. to 4 p.m.) being short, "I had a good deal of time for private study". It was a hard task he had set himself but he won through and the 1870 pass list for the degree of B.Sc. in the University of London contains the name of John Ambrose Fleming in the first division.

He now bade farewell to the Stock Exchange, though he never regretted the time he had spent there, for the knowledge of business methods thus acquired proved very useful in his later consulting work. Through the good offices of Frankland the chemist, he obtained the post of Science Master at Rossall, a public school on the Fylde coast between Blackpool and Fleetwood. Here he taught with characteristic energy and success for a year and a half and, equally characteristically, determined that he had "in some senses more need to be still a student rather than a teacher". He resigned accordingly in the autumn of 1872 and immediately entered Frankland's laboratory at the Royal College of Science in Exhibition Road, where at that time Frankland professed chemistry, Guthrie physics and Huxley biology. He became research assistant to Frankland and took a share in the teaching in the department. It was during this period that he read the paper, previously mentioned, at the inaugural meeting of the Physical Society.

In 1874 he was appointed science master at Cheltenham College on the civil and military side and was now embarked on what for many would be a life's work; but Fleming was by nature a student and an engineer. While at Cheltenham he read with care Faraday's *Experimental Researches* and, possessing a fairly strong electromagnet, repeated on a small scale Faraday's abortive experiment to detect electric currents induced in a river flowing across the earth's magnetic field. Fleming sent a stream of acidulated water down a tube placed between the poles of the magnet and carrying, fixed to opposite sides by wires sealed through the tube, two strips of platinum foil. When the magnet was energized, the strips exhibited the characteristic polarization effect. Fleming described this result in a paper read before the British Association at its Bristol meeting in 1875.

But Fleming had also been reading Maxwell's recently published treatise, and, coming to the conclusion that his mathematical equipment was defective, resigned from his safe £400 a year post and entered Cambridge in October 1877 as an Exhibitioner of St John's College.

One of Fleming's principal objects in going to Cambridge was to study and work under Maxwell, who assigned to him the task of comparing the various coils that had been constructed to represent the B.A. unit of electrical resistance and to deduce its most probable value. To that end Fleming constructed a special form of Carey Foster bridge, promptly christened by Maxwell "Fleming's banjo". He took careful and copious notes of Maxwell's lectures, and presented the bound volumes of notes to the Cavendish Laboratory on the occasion of the celebration of the centenary, in 1931, of Maxwell's birth.

In 1879 Fleming took his D.Sc. degree in London and in 1880 graduated in Cambridge by way of the Natural Science Tripos. In the same year he was appointed Demonstrator in Mechanism under Professor James Stuart and in 1882 was elected to a Fellowship at St. John's College. In the meantime he had left Cambridge, having been appointed in 1881, through the influence of Professor Stuart, to the chair of Mathematics and Physics at University College Nottingham.

But events were impending which were to make his tenure of the Nottingham post of short duration. It is difficult for those of us who saw and heard the veteran in full vigour so lately as 1938 to realize the magnitude of the changes, social and scientific, which had occurred during his lifetime and which he himself had helped so largely to shape. In the year of his birth Peel, Palmerston, Russell and Derby were outstanding political figures—Gladstone and Disraeli were, politically, promising youngsters, in the practice of war, Wellington and Nelson could have handled the armies and navies of the period without recourse to much new learning, the stage-coach had not yet vanished from the scene; the electric telegraph was a novelty.

In 1881 the storage cell was just about to become a practical proposition; Bell had invented the telephone and, realizing that the Bell telephone was an excellent receiver but a poor transmitter, Edison had invented his carbon-button transmitter, the possibilities of alternating currents were beginning to loom large, and practice in that branch of science, running ahead of theory, was to provide in the "eighties" some signal shocks and surprises for station engineers, Edison and Swan, in attempts to realize electric lighting on a commercial scale, were performing Heath-Robinsonesque experiments to produce filaments of carbon; switches, mains, cables and measuring instruments were urgently needed; patents of almost every type were filed, patent litigation, with its attendant demands for scientific witnesses (how should one, for example, define exactly a filament of carbon?) were frequent, and Fleming, who had been appointed to act as a scientific adviser to the Edison Telephone Company in 1879, was appointed as electrician to the Edison Electric Company in 1882.

In 1885 Fleming was invited to occupy the newly established Chair of Electrical Engineering at University College London. The only equipment attached to the Chair at the time of the appointment was chalk and a blackboard, but a timely grant of £150 provided funds for a gas-engine, a battery of storage cells, a photometer and some essential electrical measuring instruments. Such was the modest beginning of the Pender Chair of Electrical Engineering. Fleming resigned in 1926, after forty-one years tenure of the Chair. During that period he had made the Pender Laboratory world-famous; upwards of two thousand students had passed through his hands, many of whom achieved great distinction in after-life; a number of highly important papers were published by the Pender Professor, including an early paper read before the Institution of Electrical

Engineers in 1885 on the necessity for a national standardizing institution for electrical instruments—a paper which foreshadowed the establishing of that great organization, the National Physical Laboratory. Later came the results of measurements made on alternating-current transformers which cleared away many difficulties, pioneering researches on electric-lamp photometry, and a classic series of papers published in collaboration with Sir James Dewar on electrical properties of matter at very low temperatures.

All this research and teaching was diversified by a large amount of consulting work, the most important of which arose from his appointment as consultant to the Marconi Wireless Telegraph Company in 1899.

At that time wireless telegraphy was rapidly approaching the commercial stage; Marconi had transmitted messages in Morse between the Isle of Wight and Bournemouth; and in September 1899, at the meeting of the British Association at Dover, Fleming gave a brilliant evening address at which wireless messages were exchanged between the British Association and the French Association then meeting at Boulogne. Obviously, as with the corresponding developments in the electric telegraph half a century earlier, men's thoughts were turned to the possibility of uniting the Eastern and Western hemispheres, and one of Fleming's early tasks in his new position was the designing and equipping of a transmitting station for that purpose at Poldhu in Cornwall. The power required must have been assessed by the exercise of what the Greeks termed "happy guessing", but however that may be, a twenty-five h.p. engine, and an alternator and transformer capable of supplying current at twenty thousand volts, were installed, aërials, switches, jiggers and condensers were designed on an engineering scale; and, despite many difficulties and delays, in which stormy weather played no small part, the station was put into working order. In November 1901 Marconi sailed for Newfoundland, and on 12 December of that year—a red-letter day in the history of electrical science—he received the signals S-S-S --- transmitted from Poldhu.

Important as was Fleming's share in this triumph, his outstanding contribution to wireless telegraphy was to be made some three years later. As far back as August 1885 he had read a paper before the Physical Society in which he had noted the presence of a clear white line, or line of no deposit, on the inside of the bulb of a blackening carbon filament lamp and in the plane of the filament. He called this a molecular shadow, and gave the obvious interpretation of the phenomenon. In the previous year Edison had observed the "Edison effect" in a carbon filament lamp. He had fixed a metal plate between the hairpin bend of the filament of a lamp and had discovered that when the filament was heated by a current, a galvanoscope attached between one terminal of the filament and the plate indicated a slight current, when the attachment was made between the plate and the other terminal of the filament no current was observed. Edison gave no explanation of the phenomenon and made no notable application of it. Fleming carried out an elaborate series of experiments on cylinders and plates of various kinds sealed into filament lamps, and concluded that the heated filament was emitting negatively electrified particles, which he assumed to be carbon atoms, and that the space between filament and cylinder was a one-way street for electricity.

In the early years of the twentieth century the problem of the production of a trustworthy rectifier for feeble high-frequency oscillations was an urgent one, and in October 1904 it struck Fleming that his old experiments had provided a solution to the problem. Investigation showed that this was indeed true, and on 16 November 1904 he applied for a British patent.

Such, in very brief outline, was the origin of the Fleming valve.

Fleming had unusually high gifts as a lecturer, and his services were in constant demand over a long period of years by various educational institutions. He served for many years as a lecturer under the Gilchrist Educational Trust; he delivered eight Friday evening lectures at the Royal Institution, and between 1894 and 1922 he gave the Christmas lectures at the Royal Institution four times; his public lectures were a feature of many of the annual meetings of the British Association; and he gave the Cantor series of lectures for the Royal Society of Arts no less than eight times. His last public lecture was given before that Society in 1925 on "William Sturgeon and the Electromagnet."

His books, too, exerted a profound influence on the teaching of the period, and thousands of students of many generations have read and re-read his treatises on the transformer and on wireless telegraphy. It is a notable fact that the veteran occupied the last years of his long life in composing a series of four text-books for the use of students of engineering.

It is surely a feat unique in the history of education that a nonagenarian should produce books which should be useful to students of these later-day developments in physical science. True, there are a few slips here and there, but the old Fleming touches are still to be discerned. terseness and clarity, not a wasted word, and an essentially practical outlook

Many and well-deserved honours fell to his lot. In 1892 he was elected Fellow of the Royal Society, and he was awarded the Hughes Medal of the Society in 1910. In 1921 he was awarded the Albert Medal of the Royal Society of Arts. The Institution of Electrical Engineers gave him their Faraday Medal in 1928. In 1930 the Physical Society awarded him the Duddell Medal. In 1935 he received the Kelvin Medal from the Institution of Civil Engineers, and the Franklin Medal of the Franklin Institute. In 1929 he received the honour of knighthood. After his retirement from teaching he lived at Sidmouth in Devonshire, and in 1933 he married Miss Olive Franks. Active and interested in outside affairs to the last, he died on 19 April 1945.

There is nothing for tears, nothing for sorrow, in the passing of a life so honoured, so prolonged, so full of work which has left its mark on the scientific and social life of our era; yet, in taking an affectionate leave of the last survivor of their Founder Fellows, the Physical Society cannot help but regret that he was not spared to see the centenary of his birth and the 75th anniversary of the day on which he inaugurated the work of the Society by reading before it the first paper in their *Proceedings*.

A. F.

SIDNEY SKINNER

SIDNEY SKINNER (23 August 1863 to 5 November 1944) was educated at Dulwich College, whence he proceeded to Christ's College, Cambridge, with a scholarship in 1882. In 1884 he took the Natural Science Tripos, obtaining a second class in Part I, and a First in Part II in 1886. He stayed on at Cambridge, being appointed a lecturer both at the Cavendish Laboratory and at Clare College. He was also an assistant demonstrator at the Laboratory. In 1890 he was appointed Demonstrator, assisting Wilberforce (afterwards Professor at Liverpool) with the senior students. Searle was appointed a demonstrator at the same time and had charge of the first-year students. Other demonstrators with him were Fitzpatrick (afterwards President of Queens') and Glazebrook (later Sir Richard).

In 1904 he became Principal of the South-Western Polytechnic at Chelsea, London. At that time this Polytechnic—as all the London ones—taught a large variety of subjects, but it gradually became more specialized, and under the influence of Skinner it concentrated upon preparation for the Internal Degrees of the London University in Science, being especially strong on the biological side. The only other activities of the Polytechnic were a school of pharmacy, a good Art School and a school for preparing women to become teachers of physical training.

In 1906 he went with the expedition to Spain to see the eclipse, and there met and married Marian Field Michaelis, who had also gone there from Canada. She was daughter of Major Michaelis. Mrs Skinner was a very live woman who took a keen interest in science; both she and her husband were regular attendants at the Friday evening lectures at the Royal Institution, of which Institution he was for some years a Visitor.

They had three children, the youngest a son. Both Mrs Skinner and this son are dead.

On his retirement from the Polytechnic, in 1928, he went to live at Letchworth. There he collected a group of enthusiasts who held informal meetings and discussions on scientific subjects and were facetiously referred to by their friends as "the Wise Men."

He was a Fellow of the Institute of Physics, of the Physical Society and of the Cambridge Philosophical Society. He published nearly forty papers on a variety of subjects, the most important being on *The Clark Cell as a Standard of E.M.F.*, on *Lubrication*, and on *Dew, Dew-ponds and Humidity*.

His greatest interest was in climbing. From 1892 to 1907 he spent nearly every summer vacation in Switzerland. Two of his papers were connected with this recreation, *On the Slipperiness of Ice* and on the *Minute Structure of the Surface Ice of Glaciers*.

Skinner was a quiet conscientious worker who never cared to be in the limelight, but he had a wide circle of very real friends who appreciated his genuine ability and kindness.

R. S. C.

LIONEL ROBERT WILBERFORCE

BORN at Munich on 18 April 1861, Lionel Robert Wilberforce died on 1 April 1944 in his 83rd year. He was a direct descendant of William Wilberforce, pioneer in the abolition of the slave trade.

He was elected a life Fellow of the Physical Society in 1894 and for a period he served on the Council—his membership of the Society covered a period of almost exactly 50 years.

Educated at the London International College, Isleworth, and Trinity College, Cambridge, he obtained a first-class honours in the Natural Science Tripos and was 13th wrangler. He served as demonstrator and lecturer in Physics under J. J. Thomson from 1887 to 1900, when he was appointed Lyon Jones Professor of Physics at Liverpool University in succession to Sir Oliver Lodge. He was almost entirely responsible for the planning of the George Holt Physics Laboratories and took a great pride in the work.

Wilberforce was well grounded in the so-called "classical" physics, but took little more than a tolerant interest in the "modern" physics—quantum theory and atomic physics. As a teacher of physics he was unique. His lectures were illustrated by a plentiful display of experiments which he had specially designed and carefully rehearsed. His obvious enjoyment in demonstrating a point by a successful, and in many cases original, experiment spread to his audience.

One of his favourite experimental lectures was that relating to the principles of the gyroscope. This was always a great success and was very popular. His running commentary whilst manipulating a "diabolo" is not likely to be forgotten by his audiences.

His opinions on fundamental classical physics were sound and reliable, but he fought shy of expressing views on "modern" atomic physics. His work on surface tension, viscosity and allied problems is well known from his publications, but much of his work on these and other subjects was not published. He developed a beautiful ripple tank to demonstrate wave phenomena such as diffraction, reflection, refraction, interference, etc. I remember seeing this for the first time in 1913 and was much impressed by its possibilities in demonstrating acoustic and optical phenomena. His keen interest in instrument design is illustrated in his paper "Kinematic supports and clamps", read before the Physical Society in November 1932.

There is no doubt, however, that he will best be remembered by his students as a great teacher of physics.

He appreciated a good joke and was fond of relating his experiences in examining students. I remember his enjoyment in telling of a student he was examining in elementary practical physics at Cambridge. The subject was the spirit-level. The bubble moved to the right, the student concluding that the right-hand end of the table was higher than the left. On reversing the spirit-level, the bubble moved to the left. The student looked suspiciously at the spirit-level, then at Wilberforce, and said, after a pause, "Why, the bally thing's cockeye", a correct, if somewhat unscientific, conclusion.

Wilberforce was a member of the Alpine Club and in his younger days did a good deal of climbing in the Alps. He was very fond of skating on ice and on rollers. Such vigorous exercise, which he kept up until his later years, no doubt contributed to his rosy complexion and general good health.

It was a great pleasure to have known him. His pleasant disposition, sound judgment and success as a teacher of physics are likely to be remembered by all his many students and friends.

A. B. WOOD.

JAMES YOUNG

JAMES YOUNG, who died at the age of 47 in May 1945, after a year's unflinching fight against leukaemia, had been a physics lecturer in Birmingham University since 1920. A Scot from Strathaven, his student years had been spent in London at the Royal College of Science, and when Dr. S. W. J. Smith, F.R.S., left that College to succeed Poynting in the Birmingham professorship, he shrewdly selected Young as one of the two recent graduates of outstanding promise to accompany him.

Young possessed a combination of gifts rare in contemporary physics. He took aesthetic delight in precision, both of numerical accuracy and of logical proof, and regarded with tolerant astonishment anyone's liking for the stages in a subject which only yield

rough exploitation or ambiguous conclusion. Nevertheless he was outstandingly successful at teaching elementary students whose preoccupations would seldom dispose them to any comparable rigour, for he had the personality to command respect from those who could not approach his own intellectual ideals, and he was devoid of arrogance or contempt. He would, in fact, give limitless care to any honest enquirer, student or colleague or stranger.

This quality of the intellectual aristocrat often made Young seem lonely among any group who were advancing more rapidly in subjects less susceptible of certitude or minute precision, and he was not a very communicative colleague. Perhaps few beyond the writer heard him explain or confess that his real enthusiasm was for the geometrical in physics. This constituted him a referee of unrivalled critical judgment, rather than an explorer of new grounds. For instance, he employed his genius for three-dimensional visual imagery in elucidating complex crystalline structures, and his X-ray contributions to Prof. S. W. J. Smith's researches on nickel iron and to Prof. W. N. Haworth's new carbohydrates were notable. The same skill made him in war-time the expert on polar diagrams for Prof. M. L. Oliphant's team working at radar projection. But he had no taste for pushing into the crowd who throng the more sensational avenues to scientific eminence, and he probably spent far more time on measuring photographs of lunar craters or extracting the final decimal place in timing an occultation than on any more obviously rewarding topic.

In advanced expositions, including very individual and physical treatment of higher mechanics, trigonometrical astronomy, and Lorentz-Einstein relativity, he bequeathed to students something of his rare gifts for seeing geometry in physics. Their notebooks may well have a clarity unsurpassed in all the library of treatises which Young disdained to copy and was too self-effacing to rival in print.

MARTIN JOHNSON.

REVIEWS OF BOOKS

Practical Optics, by B. K. JOHNSON. Pp. viii+189. (London: Hatton Press, 1945.) 15s. net.

The greater precision that has been increasingly asked for in the engineering industry has resulted in the introduction of new instruments both for the measurement of dimensions to finer limits and for the closer examination of the finish. These instruments have largely made use of optical elements and devices. Thus, in addition to the older users of optical instruments—microscopists, photographers and astronomers—there has now been added a new type of user, the engineer. All these users, both old and new, want to know enough about the optics of these instruments to be able to select the instrument that will best suit their purpose and to enable them to use it intelligently, but at the same time they do not want to spend time learning more of optics than they are obliged to do, for their main interest is elsewhere. For them, therefore, a book that will just give them this minimum knowledge is a valuable time-saver. Such a book is the one under review.

It begins with some simple demonstrations which such a user should, if possible, repeat for himself, as he will find that he learns far more easily how his instruments behave from these simple experiments than he can do from merely reading about them. The principles upon which the several instruments depend are clearly explained, in every case being closely related to practical applications.

Owing to the aberrations which are present in all optical systems, every instrument is necessarily a compromise, and sacrifices some features in order to attain the greatest excellence in those which are of the greatest importance for its particular purpose. The telescope, for instance, in order to obtain a very high definition in the centre of the field, is content with an angular field which, in comparison with that of a photographic lens, is exceedingly small, and with a numerical aperture which in comparison with a micrographic object is again exceedingly small. Thus, before selecting an optical system for any particular purpose—especially if this is a new departure—it is necessary to know what are the special requirements that the system must satisfy, and the explanations given in this book will greatly assist in this selection.

We have noted the following printer's errors. The caption to figure 90 *a* should read "Measurement of Numerical Aperture", and that of figure 127 should read "Cooke Anastigmat". In the table at the foot of page 109, the last column should read 38.6, 37.0, 35.7.

Théorie des Oscillateurs, by YVES ROCARD. Pp. 220. (Paris: Editions de la Revue Scientifique, 1941). Price not stated.

M. Rocard believes that too many students of physics think that all oscillations are fundamentally similar to those of a simple pendulum, and he has written a pleasant monograph to correct this over-simplified view. The whole tone of the book is critical and, if one may be allowed to say so, refreshingly old-fashioned, old-fashioned in the sense that it is obviously written to tell the reader about oscillations and not, as are so many books on mathematical physics, written with the main object of introducing the reader to new mathematical methods.

The first chapters of the book are a critical introduction and commentary on the series of papers published by van der Pohl in the *Philosophical Magazine* on relaxation oscillators. The treatment given is elementary and straightforward and has the merit of working specific problems through in detail. In particular, the analysis of coupled oscillators, in which the causes of frequency hysteresis are explained, and the conditions for the locking of frequency-multiplying or dividing circuits obtained, will be helpful to those who have to cope with the vagaries of multivibrators. Equations such as

$$\ddot{r} - 2\omega \left[1 - r^2 - \frac{\dot{r}^2}{\omega^2} \right] r + \omega^2 r = 0,$$

which have strictly harmonic solutions, are used to poke gentle fun at some of the qualitative generalizations which are sometimes used in explaining the action of this type of oscillator. Indeed, throughout the book the tendency is to consider specific differential equations which can readily be seen to govern the motion of a simple (usually mechanical) system, and to avoid generalities.

In this way the reader is naturally introduced to systems governed by equations of the type

$$a_0 \frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} \dots + a_n y = 0$$

Routh's criterion for the instability of the system are introduced and their application explained in several examples based on mechanical models. This is an excellent chapter, and much ingenuity has been exercised in devising the rather artificial models.

More natural examples are provided in the following chapter, which deals with the differential equations which have been used in an attempt to analyse the economic oscillations of the trade cycle. The author is somewhat critical of this approach on the ground, which he justifies by numerous examples, that the predicted oscillations are unduly sensitive to small changes in the ideal mechanisms used to represent the complexities of economic life. Kalecki's equation and others which have been proposed are characterized by a period of gestation or lag between economic cause and industrial effect.

M. Rocard makes a positive contribution to the subject by pointing out that not only is the assumption of an exact value τ for the time lag a very artificial one, but also one which greatly enhances the difficulty of solving the equations. The much more plausible assumption that a lag τ has a probability given by some such function as

$$\left(e^{-\frac{\tau}{\tau_0}} \right)^n e^{-n\tau/\tau_0}$$

gives rise to equations which are much more tractable.

For example, a simple relation of the type

$$y(t-\tau) + C\dot{y} = 0$$

based upon a definite period of gestation τ becomes (for $n=1$)

$$A \int_0^\infty \tau e^{-\tau/\tau_0} y(t-\tau) d\tau + \dot{y} = 0,$$

which readily reduces by the substitution $y = e^{pt}$ to the form

$$p^2 + 2\gamma p^2 + \gamma^2 p + A = 0,$$

to which Routh's criteria can be applied directly.

The remaining chapters deal with waves on lines, the hunting of coupled generators, organ pipes, and the flutter of aeroplane wings

While these subjects are interesting in themselves, they are less suitable for a purely mathematical treatment, and do not add to the refreshing Gallic unity of the previous chapters. The section on lines, while adequate for 1940, is not of great interest to present-day students. The book was written during the fall of France, and the author asks the indulgence of his readers for the absence of references. The typography is good and the paper better than British readers have seen for some time. The absence of an index, which is partly compensated for by a full table of contents, shows that insularity occurs in the Ile de la Cité as well as at home. On the whole a book which the physicist with a taste for elementary mathematics would do well to buy.

C. H. C.

Life and Work of John Tyndall, by Professor A. S. EVE, F.R.S., and C. H. CREASEY. Pp. xxxii + 404. (London: Macmillan and Co., Ltd., 1945.) 21s. net.

We have in this book by Professor Eve and Mr C. H. Creasey a fairly complete story of the life of John Tyndall, from his boyhood at Leighlin Bridge, Co. Carlow, Ireland, where he was born on 2 August 1820, to his death on 4 December 1893. The development of the poor country lad into the head of the Royal Institution is an interesting human story. Tyndall was fortunate in that the master of the National School at Ballinabranagh, which he attended as a youth, was capable of giving him a sound training in mathematics and the rudiments of physical science, as well as in the English language.

Tyndall's first appointment was in the Ordnance Survey Office at Youghall, Co. Cork, and later he was transferred to Preston, being dismissed from the service after three and a half years for joining with 18 other assistants in an appeal against their conditions of work. After an interlude as a surveyor on new railway schemes, Tyndall obtained in 1847 a post as a master at Queenwood College, where he had Edward Frankland as a colleague. A year later Tyndall and Frankland left Queenwood together to go to the University of Marburg, where Tyndall graduated as Ph.D. after two years. While at Marburg he studied under Bunsen. One of Bunsen's assistants was Dr. Heinrich Debus, later Professor of Chemistry at the Royal Naval College, Greenwich, and one of Tyndall's closest friends. Professor Knoblauch had then just arrived at Marburg, from Berlin, and from him Tyndall got his first stimulus to study magnetism.

On his return to England he found some difficulty in obtaining a congenial post. T. H. Huxley were unsuccessful in their applications for professorships at University of Toronto in the summer of 1851 and at the University of Sydney a few months later. During this rather anxious period Tyndall was actively engaged in various researches, and was elected Fellow of the Royal Society in June 1852. In the following year he was appointed Professor of Natural Philosophy at the Royal Institution, where Faraday was Resident Professor. Faraday and Tyndall remained always on terms of the closest friendship in spite of the fact that they held different opinions on a number of scientific questions. On Faraday's death in 1867, Tyndall succeeded him as Resident Professor in the Royal Institution, and it was at the Royal Institution that Tyndall's active scientific life was passed. From the autumn of 1859 until the summer of 1866, Tyndall also held the Professorship of Physics at the School of Mines.

Tyndall was an eminently successful lecturer, and the books in which he published the content of his series of lectures are still very readable today, and have inspired many a young man with the desire to follow a scientific career. He devoted the greatest care to devising good experiments to illustrate his lectures, and the technique of the physical laboratory owed much to his ingenuity. His best known experimental investigations were those in diamagnetism, radiation and absorption by gases, and his very thorough demonstration that the moulds which Bastian had found to originate in infusions of hay, meat, fish, etc., were due to germs in the atmosphere, and were only produced when the infusions were exposed to the atmosphere. Tyndall succeeded in completely demolishing Bastian's ideas concerning the spontaneous generation of life. He also showed that the blue colour of the sky is due to scattering of light, though he ascribed the scattering to small particles and droplets suspended in the atmosphere, and it was left to the late Lord Rayleigh to demonstrate that it is the molecules of air themselves which are effective in producing the blue colour.

Tyndall derived from his father a tendency to religious emotion and controversy, and all his life he was apt to be drawn into controversies on scientific topics of all kind. He was in no sense lacking in courage, and had no hesitation in expressing his opinions even when they were unpopular. All his life he was fascinated by glaciers, and his studies of the motion of glaciers, which incidentally led him into controversy with J. D. Forbes, were among the earliest serious scientific studies of Alpine phenomena. Tyndall also devoted a great deal of energy to championing the claims for priority of scientific men whom he regarded as neglected. Thus he strongly championed Mayer's claim to have first put forward the principle of the conservation of energy.

From this book the reader receives an impression that controversy filled quite a large part of Tyndall's life. But at the same time it is made clear that Tyndall had a gift for warm friendship with men of the widest variety of interest—Faraday, T. H. Huxley, Charles Darwin, T. A. Hirst, Thomas Carlyle, Alfred Tennyson, Hooker, Lubbock. Tyndall must, in fact, have been a great social success.

It cannot be claimed that Tyndall had the spark of divine fire which Faraday had, but he was a sound and skilful experimenter, with a marked gift for exposition. He had no outstanding discovery to his credit, but he discovered many new facts, and explained many things which had hitherto been obscure. His books, which are very largely based on lectures given at the Royal Institution, formed a considerable contribution to education in the science of physics. His experimental technique was frequently far in advance of that of his contemporaries, and his breadth of interest brought him into touch with many of the vital problems of his day.

This biography of Tyndall is an important contribution to scientific history. It has a somewhat uneven style, partly due to the fact that a first draft was completed by the late Professor A. S. Eve just before his death, and was revised by Mr. Creasey. Material for an authoritative biography of John Tyndall had been collected by his wife, who survived him by nearly 47 years. Mrs. Tyndall had from time to time endeavoured herself to write the biography, and had at first endeavoured to control any effort made by others to complete the work. The unevenness of style is to be explained by the vicissitudes of authorship. It need not, however, deter anyone interested, either in the science of physics or in its history, from reading the story of a full and busy life, lived at the centre of English scientific activity, among those giants of the Victorian age—Faraday, Kelvin, Maxwell, Rayleigh, Stokes—whose work was to revitalize science and to lay the solid foundation on which later generations have built.

D. B.

An Introduction to X-ray Metallurgy, by A. TAYLOR. Pp. xi+400. (London: Chapman and Hall, Ltd., 1945.) 36s. net.

X-ray diffraction is now used extensively for investigating the structure of all classes of solids. In metallurgy particularly it has led to results of considerable practical interest. This book summarizes a great deal of material which hitherto has been available only in original scattered papers, and presents a timely and logical account of modern x-ray practice and its contributions to the study of metals and allied substances.

The first part of the book deals with the theory of x-ray diffraction and the practical methods of recording the diffraction patterns. The scope may be indicated by the chapter headings: X-ray generating apparatus, Space lattice, Diffraction of x rays by the crystal lattice, Experimental methods of obtaining diffraction patterns, Influence of the atomic pattern on the intensities of the x-ray reflections, and Crystal structures of the metals. The second part of the book then proceeds to the applications. The main topics, again described by the chapter headings, are: Study of thermal equilibrium diagrams, Measurement of grain size (including effects of deformation), Grain orientation, and Study of refractory materials. The book is then rounded off by a chapter on the older use of x rays in the radiography of metals.

In considering the theory, the author has confined himself to the bare essentials. He has taken the view, quite properly, that the application of x rays to the problems of metallurgy can be fairly appreciated without delving into the mysteries of Fourier synthesis and the theory of space groups; at the same time, it might be remarked in passing that such studies will not impede the x-ray metallographer in his career. Treatment of the practical techniques also is general, except for the Bradley powder-camera, which is

described in outstanding detail. It is perhaps to be recommended that a student should first familiarize himself thoroughly with one specific x-ray technique, and it might perhaps be added that if he can obtain the highest accuracy with this powder camera, he can proceed with confidence to any other method. However, for the practical metallurgist the reduction of a metal to powder, as a preliminary to x-ray examination, puts an unnecessary curb on the versatility of the x-ray method. A fuller description of a camera which will deal equally well with solid specimens of metal might be included with advantage in future editions of the book, which it is hoped will follow in due course.

The chapters on practical applications will be found very useful. The author has not confined himself to the earlier use of x rays in determining atomic arrangements; he has also given considerable space to the later applications concerned with the macro-structure of metals. Knowledge of the atomic structure is, of course, an essential background to any study of metals, but the atomic bond in metals is not the vital link that it is, for instance, in organic materials. In practice, much coarser features of structure are often of overriding importance, the metallurgist is often less concerned with the basic atomic arrangement than with larger scale factors such as the segregation of alloying elements and impurities, the dispersion of precipitates, crystallite size and orientation, internal strains, and grain boundary conditions. That is why the x rays will never supplant the microscope in metallurgical research, although, if intelligently used, they will supplement it with considerable advantage. The topics treated in this book well illustrate the range of application of x rays in this field. The metallurgist may consider some points over-emphasized, as the use perhaps of x rays in estimating coarse grain sizes, and some points under-stressed; but he will undoubtedly find that this part of the book makes interesting and suggestive reading.

The book fills its purpose in providing a guide to the subject, and, at the same time, a handy source of reference. It will be found useful by the student and general reader who wishes rapidly to see the high-lights of the subject, and to the research worker who may wish to refresh himself upon the many practical details involved in the interpretation of the diffraction patterns. A slight drawback is that the book is produced under war economy standards, and the reproduction of photographs, including some from papers by the reviewer, suffers somewhat in comparison with pre-war standards. Perhaps this may soon be remedied.

W. A. W.

INDEX

	PAGE
Aberration of a lens system, A differential method of adjusting the	302
Absorption, The influence of, on the shapes and positions of lines in Debye-Scherrer powder photographs	108
Achromatized plate-mirror systems	199
Andrade, E. N. da C., F.R.S. : President of the Physical Society.	<i>frontispiece</i>
Andrade, E. N. da C. . Presidential address, The history and future of the Physical Society	xxi
Aplanats, Flat-fielded singlet	567
Area of a search coil, Measurement of the effective	238
Aston, F. W., F.R.S.	<i>facing</i> 259
Balance, precision, Recent improvements in a	136
Ballard, W. R. ; The formation of metal-sprayed deposits , and discussion	67, 242
Band spectra of tin and lead halides	37
Band spectrum of MgO	12
Barkla, Charles Glover (Obituary notice)	249
Barrow, R. F. and Crawford, D. V. : A note on the spectrum of MgO	12
Bell, R. P. : A problem of heat conduction with spherical symmetry	45
Binnic, A. M. : A double-refraction method of detecting turbulence in liquids	390
Bishop, R. F., Hill, R. and Mott, N. F. : The theory of indentation and hardness tests	147
Brightness and temperature of a total radiator	440
Brightness of present-day dyes , and corrigendum	15, 146
Brown, F. G. : Exact addition formulae for the axial spherical aberration and curvature of field of an optical system of centred spherical surfaces	403
Burch, C. R. : Flat-fielded singlet aplanats	567
Caldin, E. F. : The relation between the brightness and temperature of a total radiator	440
Charlesby, A. : Effect of temperature on the structure of highly polymerized hydrocarbons	510
Charlesby, A. : Structure and orientation in thin films of polythene	496
Choong, S. P. : Coloration and luminescence produced by radium rays in the different varieties of quartz, and some optic properties of these varieties	49
Catterson-Smith, John Keats (Obituary notice)	254
Coloration and luminescence produced by radium rays in quartz	49
Colorimetry and the colours of total radiators, Illuminants for , and corrigenda	222, 370
Colour reception, The electrophysiological analysis of the fundamental problem of.	447
Combinations of spherical lenses to replace non-spherical refracting surfaces in optical systems	84
Compton profile for atoms from H to Ne, Theoretical shape of the	190
Corona, A physical theory of the solar	271
Cosmic rays, The geophysical aspect of	464
Coulson, C. A. , <i>see</i> Duncanson, W. E	
Craven, E. C. : A study of the comparative method of determining gaseous refractivities	97
Crawford, D. V., <i>see</i> Barrow, R. F.	
Critical potential experiment, A Franck and Hertz : demonstration	577
Cruikshank, F. D. : A system of transfer coefficients for use in the design of lens systems : I-V.	350, 362, 419, 426, 430
Cruikshank, F. D., <i>see</i> M'Aulay, A. L.	
Crystal lattices, Imperfections of, as investigated by the study of x-ray diffuse scattering	310
Crystals, Extrapolation methods in the derivation of accurate unit-cell dimensions of	160
Cuckow, F. W. : The electron microscope and its applications	564
Curvature of field and spherical aberration of an optical system of centred spherical surfaces	403

Davies, C N	Definitive equations for the fluid resistance of spheres.	259
Debye-Scherrer method, The determination of lattice parameters by the		126
Debye-Scherrer powder photographs, The influence of absorption on the shapes and positions of lines in		108
Deposits, metal-sprayed, The formation of		67, 242
Diatomic molecules, Resonance in precessional states of		32
Differential method of adjusting the aberration of a lens system		302
Diffusion in the ionosphere		519
Dimensions of crystals, Extrapolation methods in the derivation of		160
Directional loci in a magnetic field, and the locating of neutral points		294
Double-refraction method of detecting turbulence in liquids		390
Duddell Medallist, Twenty-first, F W Aston, F.R.S.		<i>facings</i> 259
Duncanson, W E and Coulson, C A	Theoretical shape of the Compton profile for atoms from H to Ne	190
Duperier, A	The geophysical aspect of cosmic rays	464
Dyes, The brightness of, and corrigendum		15, 146
Dynamic measurement of Young's modulus for short wires.		412
Eddington, Sir Arthur (Obituary notice)		244
Electron microscope and its applications		564
Electron microscopes (Discussion)		63
Electrophysiological analysis of the fundamental problem of colour reception		447
Equations, Definitive, for the fluid resistance of spheres		259
Expansion, Thermal, of graphite : I, Experimental ; II, Theoretical		477, 486
Extrapolation methods in the derivation of accurate unit-cell dimensions of crystals		160
Films of polythene, Structure and orientation in thin		496
Flat-fielded singlet aplanats		567
Fleming, John Ambrose (Obituary notice)		581
Fluid resistance of spheres, Definitive equations for the		259
Formulae, Exact addition, for the axial spherical aberration and curvature of field of an optical system of centred spherical surfaces		403
Franck and Hertz critical-potential experiment : demonstration		577
Frictionless state of aggregation		371
Gaseous refractivities, The comparative method of determining		97
Gent, W L, <i>see</i> Thomas, A M		
Geophysical aspect of cosmic rays		464
Goddard, L S	Discussion on electron microscopes	63
Goddard, L S and Klemperer, O	Discussion on electron microscopes	63
Granit, Ragnar	The electrophysiological analysis of the fundamental problem of colour reception	447
Graphite, The thermal expansion of	I, Experimental, II, Theoretical	477, 486
Grüneisen's equation for thermal expansion		209
Gunnar, A J : Imperfections of crystal lattices as investigated by the study of x-ray diffuse scattering		310
Guthrie Lecture, Twenty-ninth, by A. Duperier		464
H to Ne, Shape of the Compton profile for atoms from		190
Harding, H. G W	Illuminants for colorimetry and the colours of total radiators ; and corrigenda	222, 370
Hardness and indentation tests, The theory of		147
Heat conduction with spherical symmetry		45
Hill, R, <i>see</i> Bishop, R F		
Houghton, J L	Combination of spherical lenses to replace non-spherical refracting surfaces in optical systems	84
Howell, H. G	Resonance in precessional states of diatomic molecules	32
Howell, H G	The spectra of tin and lead hydrides	37
Hume-Rothery, W	Grüneisen's equation for thermal expansion	209

Hydrides, tin and lead, Spectra of	37
Hydrocarbons, Effect of temperature on the structure of highly polymerized	510
Illuminants for colorimetry and the colours of total radiators ; and corrigenda	222, 370
Imperfections of crystal lattices as investigated by the study of x-ray diffuse scattering	310
Improvements in a precision balance and the efficacy of rhodium plating for standard weights	136
Indentation and hardness tests, The theory of	147
Interferometer, Apparatus illustrating the Michelson stellar	579
Ionosphere, Note on diffusion in the	519
Jaeger, J. C., Note on diffusion in the ionosphere	519
Klemperer, O., <i>see</i> Goddard, L. S.	
Lattice parameters, The Debye-Scherrer method of determining	126
Lead and tin hydrides, The spectra of	37
Lens systems, A differential method of adjusting the aberration of a	302
Lens systems, A system of transfer coefficients for use in the design of : I-V.	350, 362, 419, 426, 430
Lenses, Combinations of spherical, to replace non-spherical refracting surfaces in optical systems	84
Lenses, Flat-fielded singlet aplanatic	567
Lenses, Theoretical investigation on telephoto	543
Linfoot, E. H. : Achromatized plate-mirror systems	199
Liquids, A double-refraction method of detecting turbulence in	390
Luminescence and coloration produced by radium rays in quartz	49
M'Aulay, A. L. : A transfer method for deriving the effect on the image formed by an optical system from ray changes produced at a given surface	435
M'Aulay, A. L. and Cruickshank, F. D. : A differential method of adjusting the aberration of a lens system	302
Magnetic field, Directional loci and the locating of neutral points in a	294
Manley, J. J. : Recent improvements in a precision balance and the efficacy of rhodium plating for standard weights	136
Mean scotopic visibility curve : corrigendum	66
Mechanical vibrations of known amplitude, A method for obtaining small	534
Mendelssohn, K. . The frictionless state of aggregation	371
Metal-sprayed deposits, The formation of	67, 242
MgO, Spectrum of	12
Michelson stellar interferometer, Apparatus illustrating the	579
Mikhail, H. and Yousef, Y. L. . The measurement of the effective area of a search coil	238
Mott, N. F., <i>see</i> Bishop, R. F.	
Nelson, J. B. and Riley, D. P. : An experimental investigation of extrapolation methods in the derivation of accurate unit-cell dimensions of crystals	160
Nelson, J. B. and Riley, D. P. . The thermal expansion of graphite from 15° c. to 800° c. : I. Experimental	477
Neutral points and directional loci in a magnetic field	294
Non-reflecting termination of a transmission line	90
Obituary notices	244, 581
Optic properties in the different varieties of quartz	49
Optical system, A transfer method for deriving the effect on the image formed by an, from ray changes produced at a given surface	435
Optical system of centred spherical surfaces, The axial spherical aberration and curvature of field of an	403

	PAGE
Optical system, A differential method of adjusting the aberration of an	302
Optical system, On tracing rays through an	286
Optical systems, Achromatized plate-mirror	199
Optical systems, A system of transfer coefficients for use in the design of ; I-V.	350, 362 419, 426, 430
Optical systems, Combinations of spherical lenses to replace non-spherical refracting surfaces in	84
Optics, Variational formulae in	558
Owen, D. : Directional loci in a magnetic field, and the locating of neutral points	294
Pelc, S. R. The photographic action of x rays	523
Permeation and sorption of water vapour in varnish films	324
Perry, W. E. : Wall- and salt-absorption corrections in radium-content measurements	178
Photocell, The internal resistance of the selenium rectifier	1
Photographic action of x rays	523
Physical Society, The history and future of the	xxi
Pincherle, L. The polarizing angle for reflection at the boundary between two absorbing media	56
Polarizing angle for reflection at the boundary between two absorbing media	56
Polymerized hydrocarbons, Effect of temperature on the structure of highly	510
Polythene, Structure and orientation in thin films of	496
Presidential Address, by E. N. da C. Andrade	xxi
Preston, J. S. and Smith, G. W. G. The internal resistance of the selenium rectifier photocell, with special reference to the sputtered metal film	1
Quartz, Coloration and luminescence produced by radium rays in, and some optic properties of	49
Radium-content measurements, Wall- and salt-absorption corrections in	178
Ray-tracing through an optical system	286
Recent reports and catalogues	66, 146, 258
Rectangular voltage waves from a low impedance source	60
Reflection at the boundary between two absorbing media, The polarizing angle for	56
Refractivities, gaseous, The comparative method of determining	97
Rehfish, T. J. : Rectangular voltage waves from a low impedance source	60
Resistance, fluid, of spheres, Definitive equations for	259
Resistance of the selenium rectifier photocell	1
Resonance in precessional states of diatomic molecules	32
Reviews of books	145, 255, 368, 444, 586
Rhodium plating for standard weights, The efficacy of	136
Riley, D. P., <i>see</i> Nelson, J. B.	
Riley, D. P. The thermal expansion of graphite : II. Theoretical	486
Saha, Meghnad. A physical theory of the solar corona	271
Sanders, J. H. Demonstration : A Franck and Hertz critical-potential experiment	577
Scattering, X-ray diffuse ; imperfections of crystal lattices as investigated by the study of	310
Scotopic visibility curve . corrigendum	66
Search coil, The measurement of the effective area of a	238
Selenium rectifier photocell, The internal resistance of the	1
Sinclair, H., <i>see</i> Taylor, A.	
Skinner, Sidney (Obituary notice)	584
Smith, D. H. : A method for obtaining small mechanical vibrations of known amplitude	534
Smith, D. H. The non-reflecting termination of a transmission line	90
Smith, G. W. G., <i>see</i> Preston, J. S.	
Smith, John Roderick Ennis (Obituary notice)	253
Smith, T. : On tracing rays through an optical system	286
Smith, T., <i>see</i> Stiles, W. S.	
Smith, T. : Theoretical investigation on telephoto lenses	543

	PAGE
Smith, T. Variational formulae in optics	558
Solar corona, A physical theory of the	271
Sorption and permeation of water vapour in varnish films	324
Spectra of tin and lead hydrides	37
Spectrum of MgO	12
Spheres, Definitive equations for the fluid resistance of	259
Spherical aberration and curvature of field of an optical system of centred spherical surfaces	403
Spherical symmetry, Heat conduction with	45
Sprayed metal deposits, The formation of	67, 242
State of aggregation, The frictionless	371
Stellar interferometer, Apparatus illustrating the demonstration	579
Stiles, W. S. and Smith, T. A mean scotopic visibility curve: corrigendum	66
Taylor, A. and Sinclair, H. : The determination of lattice parameters by the Debye-Scherrer method	126
Taylor, A. and Sinclair, H. : The influence of absorption on the shapes and positions of lines in Debye-Scherrer powder photographs	108
Telephoto lenses, Theoretical investigation on	543
Temperature and brightness of a total radiator	440
Temperature effect on the structure of highly polymerized hydrocarbons	510
Theory of indentation and hardness tests	147
Theory of the solar corona	271
Thermal expansion of graphite : I, Experimental ; II, Theoretical	477, 486
Thermal expansion, Gruneisen's equation for	209
Thomas, A. M. and Gent, W. L. : Permeation and sorption of water vapour in varnish films	324
Thomas Young Oration, Fourteenth, by Ragnar Granit	447
Tin and lead hydrides, The spectra of	37
Total radiator, the relation between The brightness and temperature of a	440
Total radiators, Illuminants for colorimetry and the colours of ; and corrigenda	222, 370
Transfer coefficients for use in the design of lens systems ; I-V	350, 362, 419, 426, 430
Transfer method for deriving the effect on the image formed by an optical system from ray changes produced at a given surface	435
Transmission line, The non-reflecting termination of a	90
Treanor, P. J. : Apparatus illustrating the Michelson stellar interferometer : demonstration	579
Tscherning, Marius Hans Erik (Obituary notice)	253
Turbulence in liquids, A double-refraction method of detecting	390
Unit-cell dimensions of crystals, Extrapolation methods in the accurate derivation of	160
Varnish films, Permeation and sorption of water vapour in	324
Variational formulae in optics	558
Vibrations of known amplitude, A method for obtaining small mechanical	534
Vickerstaff, T. : The brightness of present-day dyes , and corrigendum	15, 146
Visibility curve, Mean scotopic : corrigendum	66
Wall- and salt-absorption corrections in radium-content measurements	178
Waves from a low impedance source, Rectangular voltage	60
Weights, The efficacy of rhodium plating for standard	136
Wilberforce, L. R. (Obituary notice)	585
X-ray diffuse scattering, Imperfections of crystal lattices as investigated by the study of	310
X rays, The photographic action of	523
Young, James (Obituary notice)	585
Young's modulus for short wires, Dynamic measurement of	412
Young, V. L. : Dynamic measurement of Young's modulus for short wires	412

INDEX TO REVIEWS OF BOOKS

	PAGE
Acoustics Committee of the Building Research Board of the Department of Scientific and Industrial Research <i>Sound Insulation and Acoustics</i>	258
Brandenburg, H <i>Sechsstellige trigonometrische Tafeln</i>	370, 446
Carter, G W : <i>The Simple Calculation of Electrical Transients</i>	145
Creasey C H , <i>see</i> Eve, A S	
Dorsey, N. Ernest <i>The Velocity of Light</i>	368
Ernde, F <i>Tables of Elementary Functions</i>	368, 446
Eve, A. S and Creasey, C H <i>Life and Work of John Tyndall</i>	588
Granier, J <i>Introduction à l'étude des champs physiques</i>	445
Heitler, W <i>The Quantum Theory of Radiation</i>	446
His Majesty's Stationery Office <i>Five-figure Logarithm Tables</i>	145
Industrial Radiology Group of the Institute of Physics <i>Handbook of Industrial Radiology</i>	257
Johnson, B K : <i>Practical Optics</i>	586
Lyons, Sir Henry <i>The Royal Society, 1660-1940 I History of its Administration under the Charters</i>	255
Manley, R. G <i>Waveform Analysis</i>	444
Rocard, Yves <i>Théorie des Oscillateurs</i>	587
Schrodinger, Erwin : <i>What is Life?</i>	146
Taylor, A <i>An Introduction to X-ray Metallurgy</i>	589
Wright, W D <i>The Measurement of Colour</i>	368
Yarwood, T M <i>School Physics</i>	370

